

А. Я. ХИНЧИН

РАБОТЫ
ПО МАТЕМАТИЧЕСКОЙ
ТЕОРИИ
МАССОВОГО
ОБСЛУЖИВАНИЯ

Под редакцией Б. В. ГНЕДЕНКО



ГОСУДАРСТВЕННОЕ ИЗДАТЕЛЬСТВО
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
МОСКВА 1963

АННОТАЦИЯ

Теория массового обслуживания — важная ветвь современной теории вероятностей, развившаяся в последние годы. Эта теория может быть использована для наиболее экономного проектирования любых систем, предназначенных для удовлетворения массового потока каких-либо заявок случайного характера (например, телефонных станций, различных устройств для сбора и обработки информации и т. д.). В то же время проблемы, требующие применения тех же математических методов, возникают при автоматизации производства, организации транспорта, связи и снабжения, в военном деле.

Настоящая книга составлена из работ выдающегося советского математика, которые в своей совокупности представляют прекрасно написанное введение в изучение теории массового обслуживания. Для понимания книги необходимо владеть курсом математического анализа в объеме вузовской программы и основными понятиями теории вероятностей.

Книга рассчитана на научных работников, аспирантов и студентов старших курсов в области математики, различных отраслей техники и экономики.

СОДЕРЖАНИЕ

Предисловие редактора	4
Математические методы теории массового обслуживания	7
Математическая теория стационарной очереди	149
О среднем времени простоя стаиков	164
Потоки случайных событий без последействия	170
О пуассоновских потоках случайных событий	190
О формулах Эрланга в теории массового обслуживания	199
Теория спаренных аппаратов	209
О некоторых постановках задач и результатах теории массового обслуживания (Б. В. Гнеденко)	221

ПРЕДИСЛОВИЕ РЕДАКТОРА

В настоящей книге собраны исследования выдающегося советского математика Александра Яковлевича Хинчина (1894—1959), относящиеся к той области теории вероятностей, которую он назвал теорией массового обслуживания.

Интерес к задачам теории массового обслуживания возник у Александра Яковлевича еще в начале тридцатых годов, когда он в качестве депутата Моссовета принял участие в работе секции связи. Для Москвы тогда был период перехода к автоматическим телефонным станциям. Многочисленные научные проблемы, возникавшие при этом, увлекли А. Я. Хинчина; в результате общественное поручение наполнилось и значительным научным содержанием. Помимо работ, возникших в результате личного контакта с крупнейшими советскими представителями науки о телефонной связи, а также изучения существовавшей в ту пору литературы, Александр Яковлевич выполнил также ряд оригинальных исследований по просьбе московского телефонного узла. Связи со специалистами телефонного дела практически не прерывались буквально до последних лет жизни Хинчина. Однако творческая работа в этом направлении разбилась на два периода, отделенных друг от друга почти двадцатью годами.

В первый период активной творческой деятельности в области задач теории массового обслуживания А. Я. Хинчин опубликовал лишь две статьи: «Математическая теория стационарной очереди» (Матем. сб. 39, 1932, стр. 73—84) и «О среднем времени простоя станков» (там же 40, 1933,

стр. 119—123). В рукописях А. Я. Хинчина мне удалось найти литературно оформленную статью «Теория спаренных аппаратов», которая, как мне известно, нигде не была опубликована. С научной точки зрения особый интерес представляла первая из названных работ, так как в ней, помимо решения остро поставленной задачи вычисления распределения длительности ожидания для случая простейшего потока вызовов, поступающих на один прибор, обслуживаемых с очередью для произвольного распределения времени обслуживания, был впервые использован весьма перспективный метод. Этот метод впоследствии получил наименование «метода вложенных цепей Маркова» (см., например, работу Д. Кендалла «Стохастические процессы, встречающиеся в теории очередей, и их анализ методом вложенных цепей Маркова». Сборник переводов «Математика», 3:6, 1959, стр. 97—111).

Интерес к проблемам теории массового обслуживания пробудился у Александра Яковлевича вновь в 1953 г., вскоре после завершения цикла работ по использованию предельных теорем теории вероятностей для целей квантовой статистики. Этот интерес был вызван оживившимися связями с работниками телефонного дела. Консультации и курс лекций для инженеров, которые он тогда прочел, заставили его ознакомиться со значительной новой журнальной литературой. В результате наметились задачи, которые были восприняты А. Я. Хинчиным как центральные для всей теории массового обслуживания. Кратко эти задачи можно охарактеризовать как стремление изучить строение потока требований, поступающих на обслуживание. Собственно, именно этой задаче посвящена значительная часть его монографии «Математические методы теории массового обслуживания» (Труды математического института им. В. А. Стеклова, т. 49, Изд. АН СССР, М., 1955), а также две его последующие статьи: «Потоки случайных событий без последействия» (Теория вероятностей и ее применение, т. 1, вып. 1, 1956, стр. 3—18) и

«О пуассоновских потоках случайных событий» (там же, т. 1, вып. 3, 1956, стр. 320—327). В другом направлении было выполнено лишь исследование «О формулах Эрланга в теории массового обслуживания», где дано распространение известных формул Эрланга на случай простейшего входящего потока требований и произвольного распределения длительности обслуживания. Эта работа, будучи подготовлена к печати, не была передана автором, как мне известно, для опубликования.

Сейчас, когда интерес к теории массового обслуживания значительно вырос, издание собрания оригинальных исследований А. Я. Хинчина в этом направлении мне представляется своевременным. Я счел невозможным для себя изменять текст публикуемых работ и ограничился лишь исправлением заведомых опечаток, а также библиографическими замечаниями, которые помогут читателю при разыскании более поздних публикаций. Примечания редактора помечены буквами *Б. Г.*

Для того чтобы читатель составил себе представление о некоторых новых направлениях, в значительной мере связанных с прикладными исследованиями, в конце книги помещен небольшой обзор, написанный мною. В обзоре, в частности, указана основная монографическая литература в этой области, выпущенная в последние годы.

Б. В. Гнеденко

МАТЕМАТИЧЕСКИЕ МЕТОДЫ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

ПРЕДИСЛОВИЕ

В науке, практической деятельности людей и в быту каждодневно создаются такие положения, когда возникает массовый спрос на обслуживание какого-либо специального вида, причем обслуживающая организация, располагая лишь ограниченным числом обслуживающих единиц, не всегда способна немедленно удовлетворить все поступающие заявки. Примеры такой ситуации хорошо известны каждому. Очереди у магазинных и билетных касс, в буфетах, парикмахерских и т. д.; невозможность получить билет на нужный поезд из-за его переполнения; задержка в посадке самолетов, вызываемая отсутствием свободных посадочных площадок; задержка в ремонте потерпевших аварию станков из-за нехватки ремонтных бригад — все эти и многие другие аналогичные, хорошо известные примеры, несмотря на существенные различия их реального содержания, с формальной стороны очень близки друг другу. Во всех подобных случаях перед теорией встает, в сущности, одна основная задача: установить с возможной точностью взаимную зависимость между числом обслуживающих единиц и качеством обслуживания. При этом качество обслуживания в различных случаях, естественно, измеряется различными показателями. Большей частью таким показателем служит либо процент заявок, получающих отказ (процент пассажиров, не получивших билетов на данный поезд), либо среднее время ожидания начала обслуживания (очереди различного рода). Разумеется, качество обслуживания во всех случаях тем выше, чем больше число обслуживающих единиц; однако столь же очевидно, что чрезмерный рост этого числа сопряжен с излишним расходом сил и материальных средств; практически поэтому вопрос обычно ставится так, что сначала

устанавливается необходимый уровень качества обслуживания, а затем находится минимальное число обслуживающих единиц, при котором этот уровень может быть достигнут.

В задачах подобного рода почти всегда приходится учитывать влияние случайного элемента на течение изучаемого явления. Количество поступающих заявок не является, как правило, постоянным, а испытывает случайные колебания. Время обслуживания заявок в большинстве задач не является стандартным, а подвержено случайным колебаниям от одной заявки к другой. Все эти элементы случайности отнюдь не имеют характера небольших «возмущений», нарушающих собой плавный и закономерный ход явления; напротив, они составляют собой основную черту в картине изучаемых процессов. Естественно поэтому, что математическим инструментом теории массового обслуживания должны стать понятия и методы теории вероятностей — математической дисциплины, посвященной изучению закономерностей случая.

Цель настоящей книги — ознакомить читателя с основными идеями, методами и отдельными ходами мысли, господствующими в приложениях теории вероятностей к вопросам массового обслуживания. Необходимость создания монографии подобного рода уже давно ощущается как математиками, так и практическими работниками (в первую очередь связистами, в свое время вызвавшими к жизни изучаемую теорию и до сих пор остающимися ее главными потребителями). Эта потребность усиливается еще тем обстоятельством, что сколько-нибудь доступных изложений общей теории не имеется и за границей.

Само собой разумеется, что предлагаемая монография ни в какой мере не может претендовать на полноту сообщаемых сведений. Теория массового обслуживания в настоящее время разрослась очень широко, и для сколько-нибудь полного изложения хотя бы только важнейших из ее достижений понадобился бы толстый том. Однако такой цели я себе и не ставил. Моей задачей было — осветить на небольшом числе важнейших примеров общий характер, основной стиль использования вероятностных рассуждений в вопросах массового обслуживания. Исходя из этой «методологической» целевой установки, я и производил отбор материала для моей монографии. Особое внимание было при этом уделено трудам основоположника теории А. К. Эрланга, далеко не все исследования

которого известны у нас в той степени, какой они заслуживают. В значительном количестве мною введены в книгу также идеи К. Пальма — крупнейшего современного продолжателя дела Эрланга. В известной мере я учел и пробивающееся все чаще в наши дни стремление искать более элементарные методы исследования в противовес почти безраздельно господствующим со времени классических трудов Эрланга аналитическим методам, связанным с составлением систем дифференциальных, разностно-дифференциальных и интегральных уравнений. В этом направлении я обращаю внимание читателя на § 2 и 25, а также на всю главу 11. Попытка создания общей элементарной теории предпринята недавно в интересной статье Лундквиста (Ericsson Technics, 1953).

Составление этой монографии было очень сильно затруднено тем обстоятельством, что вся основная литература принадлежит перу практических специалистов и в математическом отношении является неудовлетворительной. Чтобы придать всему изложению форму, сколько-нибудь приемлемую в математическом смысле, я не мог оставить почти ни одного рассуждения в его первоначальном виде; приходилось либо существенно дополнять приводимую автором аргументацию, либо выбрасывать ее и заменять другой. В равной степени и вводимые новые понятия во многих случаях пришлось определять по-иному, так как определения, даваемые авторами, оказывались недостаточно четкими.

Чтобы помочь читателю связывать общие понятия с конкретными представлениями, я на протяжении всей книги пользуюсь терминологией, заимствованной из телефонной практики, запросы которой до сих пор остаются важнейшим стимулом развития теории массового обслуживания. Так, я говорю о «вызовах» (вместо «требований» или «заявок»), о «потерях» (вместо «отказов»), о длительности «разговора» (вместо «обслуживания») и т. п. Однако во всех случаях все сказанное относится, разумеется, к любому виду массового обслуживания при соответствующем изменении терминологии.

Книга разделена на три части, из которых первая — самая большая — посвящена изучению «потока» поступающих вызовов и вовсе не содержит вопросов обслуживания (которым посвящены две остальные части). Такое отчетливое выделение в особую часть изучения входящего потока казалось мне целесообразным не только потому, что для правильного

обслуживания надо прежде всего хорошо знать то, что обслуживаешь, хорошо знать все черты чередования поступающих заявок; другой, не менее важный повод к такому выделению состоял в том, что теория входящего потока вызовов есть не что иное, как общая теория случайного чередования однородных событий, которая находит себе широкий круг применений и в областях, ни с каким обслуживанием не связанных (например, в вопросах радиоактивного распада атомов).

Я стремился сделать книгу доступной всякому, кто владеет основными понятиями теории вероятностей и хотя бы кратким (вузовским) курсом математического анализа. Во многих случаях изложение могло бы быть сокращено, если бы я, вместо проведения полного доказательства того или иного утверждения, позволил себе сослаться на общие результаты соответствующей вероятностной теории (цепи Маркова, случайные процессы, эргодическая теория). Однако, желая сделать книгу более доступной возможно более широкому кругу читателей, я почти нигде не поддавался такого рода искушению.

В теории вероятностей принято называть законом распределения данной случайной величины ξ (неубывающую) функцию $F(x)$, выражающую собой вероятность неравенства $\xi \leq x$. В теории массового обслуживания часто оказывается более удобным называть законом распределения величины ξ (невозрастающую) функцию $\Phi(x)$, выражающую собой вероятность неравенства $\xi > x$. Мы будем, в зависимости от обстоятельств, пользоваться тем или другим пониманием закона распределения.

2 октября 1954 г.

А. Хинчин

ЧАСТЬ I

ВХОДЯЩИЙ ПОТОК ВЫЗОВОВ

Глава 1

ТЕОРИЯ ПРОСТЕЙШЕГО ПОТОКА

Общая теория потоков однородных событий должна естественно начинаться с определения основных общих понятий, связанных с такими потоками. Однако мы отложим такой общий подход до главы 2, где он будет проведен в нужной широте. Мы предпочитаем сразу ввести читателя в круг конкретных исследований, связанных с потоками некоторого простейшего типа, чтобы тем самым с первых страниц дать ему наглядное представление об основных ходах мысли и математических орудиях теории массового обслуживания, о стиле этого учения как математической дисциплины. После того как эти конкретные представления будут в достаточной мере усвоены, изучение более абстрактной общей теории уже не должно будет показаться трудным.

Добавим к этому, что тот простейший тип потока, который мы будем изучать в настоящей главе, в течение долгого времени оставался почти единственным, употреблявшимся в приложениях; лишь в сравнительно недавнее время отчетливо выяснилась необходимость изучения потоков более общего типа; впрочем и в наши дни значительное большинство приложений теории массового обслуживания (в частности, и приложений к телефонному делу) исходит еще из предположения, что поступающий поток требований (вызовов) принадлежит простейшему типу. Приложения (особенно технические) теории простейшего потока за последние десятилетия настолько расширились, что в настоящее время даже элементарные курсы теории вероятностей, как правило, включают в свою программу специальные главы, посвященные этой теории.

§ 1. Определение и постановка задачи

Простейшим мы будем называть поток однородных событий, если он обладает следующими тремя свойствами.

1° Стационарность. Каковы бы ни были $t > 0$ и целое $k \geq 0$, вероятность того, что за промежуток времени $(a, a + t)$ произойдет k событий, одна и та же для всех $a \geq 0$ (и, значит, зависит только от k и t); будем во всем дальнейшем обозначать эту вероятность через $v_k(t)$. На протяжении всей книги мы будем иметь дело только с такими потоками, в которых за конечный промежуток времени с вероятностью 1 происходит лишь конечное число событий. Мы будем, таким образом, всегда иметь $\sum_{k=0}^{\infty} v_k(t) = 1$ при любом t . Стационарность потока выражает собой неизменность его вероятностного режима во времени.

2° Отсутствие последствия. Вероятность $v_k(t)$ наступления k событий за промежуток времени $(a, a + t)$ не зависит от чередования событий до момента a ; другими словами, условная вероятность наступления k событий за промежуток времени $(a, a + t)$, вычисленная при любом предположении о чередовании событий до момента a , равна безусловной вероятности $v_k(t)$ того же события. Отсутствие последствия выражает собой взаимную независимость протеканий потока в непересекающихся между собой промежутках времени.

3° Ординарность. Пусть для данного стационарного потока $\psi(t)$ означает вероятность того, что за (где угодно расположенный) промежуток времени длины t наступит по меньшей мере два события [очевидно, $\psi(t) = 1 - v_0(t) - v_1(t) = \sum_{k=2}^{\infty} v_k(t)$]. Тогда мы имеем

$$\psi(t) = o(t) \quad (t \rightarrow 0),$$

или, что то же,

$$\frac{\psi(t)}{t} \rightarrow 0 \quad (t \rightarrow 0).$$

Как мы увидим далее, ординарность потока выражает собой практическую невозможность совмещения двух или более событий в один и тот же момент времени.

Итак, простейшим потоком однородных событий мы называем всякий стационарный ординарный поток без последствия.

Основная задача теории простейшего потока состоит в определении вида функций $v_k(t)$; иначе говоря, целью нашей будет отыскание закона распределения числа событий за промежуток времени длины t , рассматриваемого как случайная величина. При этом мы ради определенности реальной интерпретации изучаемого потока будем во всем дальнейшем предполагать, что речь идет о потоке вызовов, поступающих на некоторую телефонную установку, и в соответствии с этим называть наши однородные события «вызовами».

§ 2. Элементарное решение

Разобьем промежуток времени $(0, 1)$ на произвольное число n равных промежутков длины $1/n$. Вероятность того, что в какой-либо из этих частей не произойдет ни одного вызова, равна $v_0(1/n)$; а так как наш поток — без последствия, то

$$v_0(1) = \left[v_0\left(\frac{1}{n}\right) \right]^n,$$

или, полагая $v_0(1) = \theta$,

$$v_0\left(\frac{1}{n}\right) = \theta^{\frac{1}{n}}.$$

Если мы имеем промежуток длины k/n ($k=1, 2, \dots$), то он разбивается на k промежутков длины $\frac{1}{n}$, вследствие чего

$$v_0\left(\frac{k}{n}\right) = \left[v_0\left(\frac{1}{n}\right) \right]^k = \theta^{k/n}. \quad (2.1)$$

Пусть, наконец, t — любое положительное число и пусть натуральное число k определяется из неравенств

$$\frac{k-1}{n} < t \leq \frac{k}{n};$$

так как $v_0(t)$, очевидно, есть невозрастающая функция от t , то

$$v_0\left(\frac{k-1}{n}\right) \geq v_0(t) \geq v_0\left(\frac{k}{n}\right),$$

или, в силу (2.1),

$$\theta^{\frac{k-1}{n}} \geq v_0(t) \geq \theta^{\frac{k}{n}};$$

но при $n \rightarrow \infty$ мы имеем $\frac{k}{n} \rightarrow t$, вследствие чего крайние члены написанных неравенств стремятся к θ^t , и мы находим

$$v_0(t) = \theta^t$$

для любого $t > 0$. При этом постоянное число θ нами определено как $v_0(1)$, и следовательно, $0 \leq \theta \leq 1$. Однако случаи $\theta = 0$ и $\theta = 1$ не представляют интереса, и мы можем их не рассматривать. В самом деле, при $\theta = 1$ мы имеем $v_0(t) = 1$ при любом $t > 0$, что означает достоверное отсутствие вызовов в любом промежутке времени, т. е. отсутствие какого бы то ни было потока. Если же $\theta = 0$, то $v_0(t) = 0$ при любом $t > 0$; это означает, что вызовы с достоверностью будут получены в любом, сколь угодно малом промежутке времени; но тогда, сколь бы велико ни было k , число вызовов в любом промежутке с достоверностью будет больше, чем k ; другими словами, число вызовов в любом промежутке бесконечно с вероятностью 1; но такие потоки мы в § 1 раз навсегда исключили из рассмотрения. Итак, мы можем считать, что $0 < \theta < 1$; поэтому можно положить $\theta = e^{-\lambda}$, где λ — постоянное положительное число, и писать

$$v_0(t) = e^{-\lambda t}. \quad (2.2)$$

Отметим, что при этом выводе мы нигде не пользовались ординарностью нашего потока, так что соотношение (2.2) имеет силу для любого стационарного потока без последовательности; это замечание будет для нас важно в дальнейшем.

Теперь обращаемся к отысканию функций $v_k(t)$ при $k > 0$. Существует много различных способов решения этой задачи, и почти все они поучительны, так как заложенные в них методы позволяют решать и ряд более сложных задач. Мы начнем с наиболее элементарного способа. Будем считать t постоянным и разобьем промежуток $(0, t)$ на произвольное число $n > k$ равных частей (ячеек) длины $t/n = \delta$. Относительно расположения вызовов в этих ячейках возможны две гипотезы:

H_1 — ни в одной из n ячеек не будет более одного вызова;

H_2 — по крайней мере в одной из ячеек произойдет более одного вызова.

Мы, очевидно, имеем

$$v_k(t) = P(H_1, k) + P(H_2, k), \quad (2.3)$$

где $P(H_i, k)$ ($i=1, 2$) означает вероятность двойного события: 1) реализуется гипотеза H_i и 2) в промежутке $(0, t)$ поступает k вызовов. Очевидно, что $P\{H_1, k\}$ есть вероятность такого положения вещей, когда из наших n ячеек какие-либо k содержат по одному вызову, а остальные $n-k$ вообще вызовов не содержат, поэтому

$$P(H_1, k) = \binom{n}{k} [v_1(\delta)]^k [v_0(\delta)]^{n-k}.$$

В силу формулы (2.2) и ординарности данного потока мы имеем при $n \rightarrow \infty$ ($\delta \rightarrow 0$) и постоянном k

$$\begin{aligned} [v_0(\delta)]^{n-k} &= e^{-\lambda\delta(n-k)} = e^{-\lambda t} e^{k\lambda\delta} = e^{-\lambda t} [1 + o(1)]; \\ [v_1(\delta)]^k &= [1 - e^{-\lambda\delta} - \psi(\delta)]^k = [1 - e^{-\lambda\delta} + o(\delta)]^k = \\ &= (\lambda\delta)^k [1 + o(1)] = \frac{(\lambda t)^k}{n^k} [1 + o(1)], \end{aligned}$$

и, следовательно,

$$\begin{aligned} P(H_1, k) &= \\ &= e^{-\lambda t} \frac{(\lambda t)^k}{k!} \frac{n(n-1)\dots(n-k+1)}{n^k} [1 + o(1)] \rightarrow e^{-\lambda t} \frac{(\lambda t)^k}{k!} \end{aligned}$$

при $n \rightarrow \infty$.

С другой стороны, $P(H_2, k)$, очевидно, не превосходит вероятности гипотезы H_2 , т. е. того, что по меньшей мере одна из n ячеек содержит более одного вызова; так как для отдельной ячейки вероятность содержать более одного вызова есть $\psi(\delta)$, то поэтому

$$P(H_2, k) \leq n\psi(\delta) = t \frac{\psi(\delta)}{\delta} \rightarrow 0 \quad (n \rightarrow \infty);$$

таким образом, правая часть равенства (2.3) при $n \rightarrow \infty$ имеет пределом

$$e^{-\lambda t} \frac{(\lambda t)^k}{k!};$$

а так как левая часть (2.3) от n не зависит, то

$$v_k(t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!} \quad (k=0, 1, \dots).$$

Таким образом, для простейшего потока число вызовов в промежутке длины t распределено по закону Пуассона с параметром λt .

§ 3. Метод дифференциальных уравнений

В специальной литературе поставленная нами задача решается обычно другим методом, менее элементарным, но зато легко распространяемым на более сложные задачи. Рассмотрим теперь этот метод. Пусть t и τ — любые положительные числа и пусть $k > 0$. Пусть k_1 — число вызовов в промежутке $(0, t)$, а k_2 — в промежутке $(t, t + \tau)$. Для того чтобы в промежутке $(0, t + \tau)$ произошло k вызовов, необходимо и достаточно наступление одного из следующих двойных событий: $k_1 = k, k_2 = 0$; $k_1 = k - 1, k_2 = 1$; $k_1 = k - 2, k_2 = 2$; ...; $k_1 = 0, k_2 = k$. Но вероятности событий $k_1 = l$ и $k_2 = m$ соответственно равны $v_l(t)$ и $v_m(\tau)$; а так как эти события взаимно независимы (поток без последовательности!), то

$$v_k(t + \tau) = v_k(t)v_0(\tau) + v_{k-1}(t)v_1(\tau) + \dots + v_0(t)v_k(\tau). \quad (3.1)$$

Но при $\tau \rightarrow 0$ мы имеем:

$$\begin{aligned} v_0(\tau) &= e^{-\lambda\tau} = 1 - \lambda\tau + o(\tau); \\ v_1(\tau) &= 1 - v_0(\tau) = \lambda\tau + o(\tau); \\ \sum_{i=2}^k v_{k-i}(t)v_i(\tau) &\leq \sum_{i=2}^{\infty} v_i(\tau) = \psi(\tau) = o(\tau). \end{aligned}$$

и, следовательно, (3.1) дает

$$\begin{aligned} v_k(t + \tau) &= v_k(t)(1 - \lambda\tau) + v_{k-1}(t)\lambda\tau + o(\tau); \\ \frac{v_k(t + \tau) - v_k(t)}{\tau} &= \lambda[v_{k-1}(t) - v_k(t)] + o(1). \end{aligned}$$

Это показывает, что функция $v_k(t)$ дифференцируема при любом $t > 0$ и что

$$v'_k(t) = \lambda[v_{k-1}(t) - v_k(t)] \quad (k=1, 2, \dots); \quad (3.2)$$

если положить для общности $v_{-1}(t) \equiv 0$, то уравнение (3.2), как мы непосредственно убеждаемся из (2.2), имеет место и при $k=0$.

Таким образом, для определения искомых функций $v_k(t)$ мы получили систему линейных дифференциальных уравнений (3.2). Эта система легко решается различными методами, из которых мы рассмотрим два наиболее поучительных для дальнейшего.

А. Метод замены искомых функций

Положим

$$v_k(t) = e^{-\lambda t} u_k(t) \quad (k=0, 1, 2, \dots),$$

отсюда

$$v'_k(t) = e^{-\lambda t} [u'_k(t) - \lambda u_k(t)] \quad (k=0, 1, 2, \dots);$$

вставляя же эти выражения $v_k(t)$ и $v'_k(t)$ в уравнения (3.2), легко находим

$$u'_k(t) = \lambda u_{k-1}(t) \quad (k=0, 1, 2, \dots),$$

где по определению $u_{-1}(t) \equiv 0$. Отсюда, интегрируя, находим

$$u_k(t) - u_k(0) = \lambda \int_0^t u_{k-1}(z) dz \quad (k=0, 1, 2, \dots).$$

Очевидно, мы имеем при любом $k \geq 0$

$$v_k(0) = u_k(0);$$

но по определению функций $v_k(t)$

$$v_0(0) = 1, \quad v_k(0) = 0 \quad (k > 0);$$

поэтому и $u_0(0) = 1$, $u_k(0) = 0$ ($k=1, 2, \dots$), и мы находим при $k \geq 1$

$$u_k(t) = \lambda \int_0^t u_{k-1}(z) dz. \quad (3.3)$$

Замечая, что в силу (2.2) мы имеем по определению функций $u_k(t)$

мы по формуле (3.3) рекуррентно находим

$$\begin{aligned} u_1(t) &= \lambda t; \\ u_2(t) &= \frac{(\lambda t)^2}{2}; \\ &\dots \dots \dots; \\ u_k(t) &= \frac{(\lambda t)^k}{k!}; \\ &\dots \dots \dots, \end{aligned}$$

и, следовательно,

$$v_k(t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!},$$

т. е. получаем прежнее решение задачи.

В. Метод производящих функций

Положим

$$\sum_{k=0}^{\infty} v_k(t) x^k = \Phi(t, x);$$

ряд в левой части этого равенства во всяком случае абсолютно сходится при $|x| \leq 1$. Умножая на x^k все члены уравнения (3.2) и суммируя по k от 0 до ∞ , мы легко находим

$$\frac{\partial \Phi}{\partial t} = \lambda \sum_{k=0}^{\infty} v_{k-1}(t) x^k - \lambda \Phi = \lambda (x - 1) \Phi,$$

или

$$\frac{\partial \ln \Phi}{\partial t} = \lambda (x - 1);$$

отсюда

$$\ln \Phi(t, x) - \ln \Phi(0, x) = \lambda (x - 1) t; \quad (3.4)$$

но легко видеть, что при любом x

$$\Phi(0, x) = v_0(0) = 1,$$

поэтому (3.4) дает

$$\Phi(t, x) = e^{\lambda (x-1)t} = e^{-\lambda t} e^{\lambda t x} = e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} x^k.$$

Сопоставляя это с определением функции $\Phi(t, x)$, мы непосредственно видим, что

$$v_k(t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!} \quad (k=0, 1, 2, \dots),$$

т. е. снова приходим к прежнему решению задачи.

§ 4. Интенсивность простейшего потока

Полученные нами результаты показывают, что те три свойства (стационарность, отсутствие последствия, ординарность), которыми мы определили простейший поток, полностью характеризуют его структуру с точностью до значения параметра λ , которое может быть любым положительным числом. Два простейших потока могут отличаться друг от друга только значениями этого параметра.

Условимся обозначать в дальнейшем для любого стационарного потока через $w(t)$ вероятность того, что за промежуток времени t произойдет по меньшей мере один вызов. Очевидно, мы имеем

$$w(t) = 1 - v_0(t) = \sum_{k=1}^{\infty} v_k(t) = v_1(t) + \psi(t),$$

где $\psi(t) = \sum_{k=2}^{\infty} v_k(t)$ по-прежнему означает вероятность поступления по меньшей мере двух вызовов за промежуток времени длины t . Для простейшего потока с параметром λ $v_0(t) = e^{-\lambda t}$ и, значит, при $t \rightarrow 0$

$$w(t) = 1 - e^{-\lambda t} = \lambda t + o(t),$$

или, что то же,

$$\lim_{t \rightarrow 0} \frac{w(t)}{t} = \lambda. \quad (4.1)$$

Мы можем считать это соотношение *определением* параметра λ для данного потока. Мы узнаем в дальнейшем, что предел (4.1) существует у любого стационарного потока и определенный соотношением (4.1) параметр λ служит одной из важнейших характеристик этого потока.

Но вернемся к простейшему потоку и найдем теперь математическое ожидание числа вызовов, поступающих за

промежуток времени длины t . Оно равно

$$\sum_{k=1}^{\infty} k v_k(t) = e^{-\lambda t} \sum_{k=1}^{\infty} \frac{(\lambda t)^k}{(k-1)!} = e^{-\lambda t} \lambda t \sum_{k=1}^{\infty} \frac{(\lambda t)^{k-1}}{(k-1)!} = \lambda t,$$

так как последняя сумма, очевидно, равна $e^{\lambda t}$. Мы могли бы предвидеть этот результат и заранее, так как известно, что математическое ожидание величины, распределенной по закону Пуассона, равно параметру этого закона, т. е. в данном случае равно λt .

Математическое ожидание числа вызовов в единицу времени называют интенсивностью данного потока; мы будем обозначать эту интенсивность через μ . Как мы только что установили, для простейшего потока $\mu = \lambda$. Однако для стационарных потоков более сложной структуры это равенство не только не очевидно, но и не всегда верно; в этом вопросе мы подробно разберемся в дальнейшем. Сейчас же убедимся только, что для любого стационарного потока $\mu \geq \lambda$. В самом деле, математическое ожидание числа вызовов за время t для данного потока равно

$$\mu t = \sum_{k=1}^{\infty} k v_k(t) \geq \sum_{k=1}^{\infty} v_k(t) = w(t),$$

$$\mu \geq \frac{w(t)}{t},$$

а так как левая часть этого неравенства от t не зависит, то в силу (4.1) $\mu \geq \lambda$; разумеется, самое существование предела (4.1) для любого стационарного потока еще должно быть доказано.

Итак, интенсивность μ простейшего потока совпадает с его параметром λ ; для произвольного же стационарного потока мы пока можем утверждать только, что $\mu \geq \lambda$. При этом существование предела (4.1) для любого стационарного потока еще должно быть доказано *).

* В сущности и параметр λ может быть реально интерпретирован как «интенсивность» данного потока, так как соотношение (4.1) показывает, что вероятность поступления вызовов в промежутке бесконечно малой длины t асимптотически пропорциональна t , и коэффициентом пропорциональности служит как раз параметр λ . Можно было бы говорить о «верхней интенсивности» μ и «нижней интенсивности» λ (так как всегда $\mu \geq \lambda$). В дальнейшем (§ 11) мы узнаем, что для ординарного потока всегда $\mu = \lambda$.

§ 5. Поток с переменным параметром

В этой книге мы будем изучать почти исключительно стационарные потоки вызовов. Однако для некоторых простейших задач решение в нестационарном случае настолько легко проводится и вместе с тем имеет столь ясное практическое значение, что было бы жаль оставить его совсем без рассмотрения. В частности, в настоящем параграфе мы подвергнем изучению потоки, не обладающие стационарностью, но являющиеся, подобно простейшему потоку, ординарными потоками без последействия. Мы сейчас более точно поясним смысл этих предпосылок.

Если поток не стационарен, то вероятность получить k вызовов за промежуток времени длины τ зависит не только от τ , но и от начального момента t этого промежутка: поэтому мы будем обозначать ее через $v_k(\tau, t)$. Таким образом, $v_k(\tau, t)$ есть вероятность того, что за промежуток времени $(t, t + \tau)$ произойдет k вызовов. По аналогии со стационарным случаем мы полагаем

$$1 - v_0(\tau, t) = w(\tau, t); \quad 1 - v_0(\tau, t) - v_1(\tau, t) = \psi(\tau, t).$$

Мы будем называть исследуемый поток ординарным, если при $\tau \rightarrow 0$ и любом постоянном $t \geq 0$ имеет место соотношение

$$\frac{\psi(\tau, t)}{\tau} \rightarrow 0.$$

Далее мы должны допустить, что для любого $t \geq 0$ существует

$$\lim_{\tau \rightarrow 0} \frac{w(\tau, t)}{\tau} = \lambda(t) \quad (5.1)$$

(мгновенное значение параметра).

Исходя из этих предпосылок, мы поставим себе задачей найти выражение функций $v_k(\tau, t)$. Рассмотрим, как и прежде, сначала случай $k=0$.

Так как мы имеем дело с потоком без последействия, то при $\Delta\tau > 0$

$$v_0(\tau + \Delta\tau, t) = v_0(\tau, t) v_0(\Delta\tau, t + \tau);$$

но по предположению при $\Delta\tau \rightarrow 0$ и постоянных t, τ

$$v_0(\Delta\tau, t + \tau) = 1 - w(\Delta\tau, t + \tau) = 1 - \lambda(t + \tau)\Delta\tau + o(\Delta\tau),$$

следовательно,

$$v_0(\tau + \Delta\tau, t) - v_0(\tau, t) = -v_0(\tau, t) \lambda(t + \tau) \Delta\tau + o(\Delta\tau);$$

это после почленного деления на $\Delta\tau$ в пределе приводит к соотношению

$$\frac{\partial v_0(\tau, t)}{\partial \tau} = -\lambda(t + \tau) v_0(\tau, t) \quad (5.2)$$

(причем существование производной, очевидно, попутно доказывается); отсюда же

$$\frac{\partial \ln v_0(\tau, t)}{\partial \tau} = -\lambda(t + \tau),$$

и, следовательно,

$$\ln v_0(\tau, t) - \ln v_0(0, t) = -\int_0^\tau \lambda(t + u) du;$$

а так как $\ln v_0(0, t) = 0$, то

$$\begin{aligned} \ln v_0(\tau, t) &= -\int_0^\tau \lambda(t + u) du = -\Lambda(\tau, t); \\ v_0(\tau, t) &= e^{-\Lambda(\tau, t)}. \end{aligned}$$

В стационарном случае мы имели показателем $-\lambda\tau$; в общем случае, как мы теперь видим, нужно заменить λ величиной

$$\frac{1}{\tau} \Lambda(\tau, t) = \frac{1}{\tau} \int_0^\tau \lambda(t + u) du,$$

которую естественно рассматривать как среднее значение «мгновенного параметра» $\lambda(t)$ в промежутке $(t, t + \tau)$.

Переходя теперь к случаю $k > 0$, мы аналогично предыдущему [см. § 3] при $\Delta\tau \rightarrow 0$ и постоянных t, τ легко находим

$$v_k(\tau + \Delta\tau, t) = v_k(\tau, t) v_0(\Delta\tau, t + \tau) + \\ + v_{k-1}(\tau, t) v_1(\Delta\tau, t + \tau) + o(\Delta\tau),$$

где

$$v_0(\Delta\tau, t + \tau) = 1 - \lambda(t + \tau) \Delta\tau + o(\Delta\tau)$$

и

$$\begin{aligned} v_1(\Delta\tau, t + \tau) &= 1 - v_0(\Delta\tau, t + \tau) - \psi(\Delta\tau, t + \tau) = \\ &= \lambda(t + \tau) \Delta\tau + o(\Delta\tau), \end{aligned}$$

так что

$$v_k(\tau + \Delta\tau, t) = v_k(\tau, t) [1 - \lambda(t + \tau) \Delta\tau] + \\ + v_{k-1}(\tau, t) \lambda(t + \tau) \Delta\tau + o(\Delta\tau),$$

откуда

$$\frac{v_k(\tau + \Delta\tau, t) - v_k(\tau, t)}{\Delta\tau} = \lambda(t + \tau) [v_{k-1}(\tau, t) - v_k(\tau, t)] + o(1),$$

и следовательно, в пределе

$$\frac{\partial v_k(\tau, t)}{\partial \tau} = \lambda(t + \tau) [v_{k-1}(\tau, t) - v_k(\tau, t)]. \quad (5.3)$$

Это соотношение, доказанное нами для любого $k > 0$, остается, как показывает (5.2), верным и при $k = 0$, если положить

$$v_{-1}(\tau, t) \equiv 0.$$

Мы найдем нужное нам решение системы (5.3), применяя метод производящих функций. Положим

$$F(t, \tau, x) = \sum_{k=0}^{\infty} v_k(\tau, t) x^k.$$

Умножая все члены уравнения (5.3) на x^k и суммируя по k от 0 до ∞ , мы находим в точной аналогии с § 3

$$\frac{\partial F}{\partial \tau} = (x - 1) \lambda(t + \tau) F,$$

или

$$\frac{\partial \ln F}{\partial \tau} = (x - 1) \lambda(t + \tau),$$

откуда

$$\ln F(t, \tau, x) - \ln F(t, 0, x) = (x - 1) \int_0^{\tau} \lambda(t + u) du = \\ = (x - 1) \Lambda(\tau, t). \quad (5.4)$$

При любых x и t мы имеем

$$F(t, 0, x) = v_0(0, t) = 1;$$

поэтому (5.4) дает

$$F(t, \tau, x) = e^{(x-1) \Lambda(\tau, t)} = e^{-\Lambda(\tau, t)} e^{x \Lambda(\tau, t)} = \\ = \sum_{k=0}^{\infty} e^{-\Lambda(\tau, t)} \frac{[\Lambda(\tau, t)]^k}{k!} x^k,$$

и сопоставление с определением функции $F(t, \tau, x)$ дает

$$v_k(\tau, t) = e^{-\Lambda(\tau, t)} \frac{[\Lambda(\tau, t)]^k}{k!} \quad (k=0, 1, 2, \dots). \quad (5.5)$$

Эти формулы полностью решают поставленную задачу. Мы видим, что и для потока с переменным параметром число вызовов в промежутке $(t, t + \tau)$ подчиняется закону Пуассона; однако параметр этого закона теперь зависит не только от длины τ данного промежутка, но и от его начального момента t . В случае стационарного потока мы имели закон Пуассона с параметром $\lambda\tau$; при переходе к нестационарному случаю мы должны, как видим, заменить постоянное число λ выражением

$$\frac{\Lambda(\tau, t)}{\tau} = \frac{1}{\tau} \int_0^\tau \lambda(t+u) du,$$

т. е. средним значением $\lambda(x)$ в промежутке $t \leq x \leq t + \tau$. Обратно, если функция $\lambda(x) = \lambda$ есть постоянная величина, то, очевидно, при любом t

$$\Lambda(\tau, t) = \int_0^\tau \lambda(t+u) du = \lambda\tau$$

и формулы (5.5) переходят в решения, полученные нами в § 3 для стационарного случая.

Заметим, наконец, что число вызовов в промежутке $(t, t + \tau)$, подчиняясь закону Пуассона (5.5), имеет своим математическим ожиданием параметр этого закона, т. е. величину

$$\Lambda(\tau, t) = \int_t^{t+\tau} \lambda(u) du;$$

поэтому величину $\frac{1}{\tau} \Lambda(\tau, t)$ можно понимать как *среднюю* интенсивность нашего потока в промежутке $(t, t + \tau)$; предел же этой величины при $\tau \rightarrow 0$ есть *мгновенная* интенсивность $\mu(t)$ данного потока в момент t ; мы находим

$$\mu(t) = \lim_{\tau \rightarrow 0} \frac{1}{\tau} \Lambda(\tau, t) = \lim_{\tau \rightarrow 0} \frac{1}{\tau} \int_t^{t+\tau} \lambda(u) du = \lambda(t).$$

Таким образом, и в случае простейшего потока с переменным параметром мы имеем совпадение мгновенной интенсивности потока с мгновенным значением параметра.

Глава 2

ОБЩИЕ СВОЙСТВА СТАЦИОНАРНЫХ ПОТОКОВ *)

§ 6. Поток вызовов как случайный процесс

Как мы уже говорили в начале главы 1, потоки вызовов, с которыми мы встречаемся на практике, во многих случаях могут с достаточно хорошим приближением рассматриваться как простейшие; при изучении таких потоков поэтому обычно пользуются теми результатами, которые получены нами в главе 1. Однако в последние годы, когда усложняющаяся практика ставит перед наукой задачи все более сложные и требующие все более точного решения, встала настоятельная необходимость изучения потоков более общего типа. Непосредственных поводов для такого расширения изучаемой области имеется в основном два.

С одной стороны, статистика потоков вызовов даже в самых обычных условиях при возрастающей точности показывает, что выводы, основанные на предположении простейшего характера потока, недостаточно хорошо согласуются с опытными данными; можно без труда и теоретически предвидеть необходимость такого рода расхождений: нетрудно сообразить, например, что в действительности почти всегда следует ожидать в изучаемом потоке известного последействия и что не всегда этим последствием можно пренебрегать.

С другой стороны, имеются и такие случаи, когда изучаемый поток заведомо и принципиально отличается от простейшего; таковы все виды потоков переменной интенсивности (они не стационарны, см. § 5); таковы же и потоки, поступающие на вторую, третью, и т. д. линию «полнодоступного пучка», даже в том случае, когда на первую линию поступает простейший поток; все эти потоки обладают значитель-

*) Естественным продолжением результатов настоящей главы являются работы А. Я. Хиичина «Потоки случайных событий без последействия» и «О пуассоновских потоках случайных событий», помещенные в настоящей книге (стр. 170 и 190). — Б. Г.

ным последствием, возрастающим с номером линии, и учет этого последствия обязателен для теории (см. гл. 8). Поэтому современные исследования по теории массового обслуживания не могут ограничиться рассмотрением простейших потоков и вынуждены расширить в той или другой мере исходные предпосылки.

Переходя к исследованию потоков более общего типа, мы должны в целях строгости и недвусмысленной ясности изложения начать с точного определения основных понятий; мы не сделали этого в главе 1, так как рассматривали эту главу как вводную, имеющую целью на простейшем примере показать характерные для всей теории потоков образцы применяемых в ней математических методов.

Если мы обозначим через $x(t)$ число вызовов, поступающих за промежуток времени $(0, t)$, то для каждого фиксированного значения $t > 0$ $x(t)$ представляет собой случайную величину. При переменном t , $x(t)$ представляет собой однопараметрическое семейство случайных величин, которое называют случайным процессом или случайной функцией. Для функции $x(t)$ характерно то, что она: 1) может принимать только целые неотрицательные значения и 2) с возрастанием t никогда не убывает. График такой функции

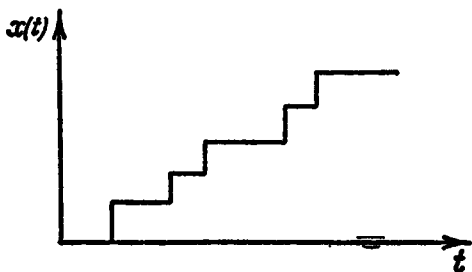


Рис. 1.

поэтому, независимо от случая, всегда имеет форму «лесенки», изображенную на рис. 1.

Для задания любого случайного процесса $x(t)$ как такового надо, чтобы для любой конечной группы положительных чисел $t_1, t_2, \dots, t_n$ был задан n -мерный закон распре-

деления вектора

$$x(t_1), x(t_2), \dots, x(t_n).$$

Если процесс $x(t)$ представляет собой поток вызовов и, следовательно, $x(t)$ может принимать только целые неотрицательные значения, то для задания этого потока как случайного процесса надо задать для каждой группы положительных $t_1, t_2, \dots, t_n$ и каждой группы целых неотрицательных чисел $k_1, k_2, \dots, k_n$ вероятность системы равенств $x(t_1) = k_1, x(t_2) = k_2, \dots, x(t_n) = k_n$; очевидно, эта вероятность может быть отличной от нуля только в том случае, если при $t_1 < t_2 < \dots < t_n$ мы имеем $k_1 \leq k_2 \leq \dots \leq k_n$. В частности ($n=1$), для любого $t > 0$ и любого целого неотрицательного k должна быть известна вероятность равенства $x(t) = k$, которую мы в главе 1 обозначали через $v_k(t)$ *). Таким образом, система функций $v_k(t)$ ($k=0, 1, 2, \dots$) входит как обязательный элемент в состав описания каждого потока вызовов. В общем случае, однако, задания этой системы функций для полной характеристики потока еще недостаточно.

Поток вызовов называется стационарным, если при $0 < t_1 < t_2 < \dots < t_n$ и при любом положительном a закон распределения вектора $x(t_i)$ ($1 \leq i \leq n$) совпадает с законом распределения вектора $x(a + t_i) - x(a)$ ($1 \leq i \leq n$); иначе говоря, закон распределения вектора $x(a + t_i) - x(a)$ ($1 \leq i \leq n$) зависит от чисел t_i , но не зависит от a . В частности ($n=1$), $v_k(t)$ для стационарного процесса означает вероятность поступления k вызовов в промежутке $(a, a + t)$, где $a \geq 0$ произвольно, т. е. в любом промежутке длины t **).

Данный поток вызовов называется потоком без последствия, если закон распределения вектора

*) Необходимо, впрочем, отметить, что в главе 1 $v_k(t)$ (в силу предположенной стационарности потока) означало вероятность поступления k вызовов в любом промежутке времени длины t , в то время как здесь речь идет лишь о промежутке $(, t)$.

**) Для читателя, знакомого с теорией случайных процессов, заметим, что определенный таким образом стационарный поток вызовов не является, конечно, стационарным случайным процессом в общепринятом смысле этого термина. В теории случайных процессов наш стационарный поток принадлежит к числу «процессов со стационарными приращениями».

$x(a+t_i) - x(a)$ ($t_i > 0$, $1 \leq i \leq n$) при любом $a \geq 0$ не зависит от значений величины $x(t)$ при каких-либо значениях $t < a$. Это определение, очевидно, в точной форме выражает то требование, чтобы случайное течение потока вызовов после какого-либо момента времени a было независимым от его течения до момента a ; но в этом и состоит отсутствие последствия в понимании теории вероятностей.

Легко видеть, что стационарный поток без последствия полностью характеризуется системой функций $v_k(t)$ ($k=0, 1, 2, \dots$), т. е. законом распределения числа вызовов, поступающих в течение (где угодно расположенного) промежутка времени длины t . В самом деле, так как система равенств

$$x(t_i) = k_i \quad (1 \leq i \leq n, 0 < t_1 < t_2 < \dots < t_n),$$

очевидно, равносильна системе равенств

$$x(t_i) - x(t_{i-1}) = k_i - k_{i-1} \quad (1 \leq i \leq n),$$

где для общности положено $t_0 = k_0 = x(t_0) = 0$, то (обозначая через $P\{\}$ вероятность события, помещенного в фигурных скобках) мы будем иметь

$$\begin{aligned} P\{x(t_i) = k_i, 1 \leq i \leq n\} = \\ = P\{x(t_i) - x(t_{i-1}) = k_i - k_{i-1}, 1 \leq i \leq n\}; \end{aligned}$$

так как промежутки (t_0, t_1) , (t_1, t_2) , $\dots$, (t_{n-1}, t_n) взаимно не перекрываются, то отсюда в силу отсутствия последствия

$$\begin{aligned} P\{x(t_i) = k_i, 1 \leq i \leq n\} = \\ = \prod_{i=1}^n P\{x(t_i) - x(t_{i-1}) = k_i - k_{i-1}\}, \end{aligned}$$

а так как в силу стационарности потока

$$\begin{aligned} P\{x(t_i) - x(t_{i-1}) = k_i - k_{i-1}\} = P\{x(t_i - t_{i-1}) = k_i - \\ - k_{i-1}\} = v_{k_i - k_{i-1}}(t_i - t_{i-1}) \quad (1 \leq i \leq n), \end{aligned}$$

то *)

$$P\{x(t_i) = k_i, 1 \leq i \leq n\} = \prod_{i=1}^n v_{k_i - k_{i-1}}(t_i - t_{i-1}).$$

*) Здесь и в дальнейшем мы, разумеется, полагаем $v_r(t) = 0$ для любого отрицательного индекса r .

Это показывает, что заданием системы функций $v_k(t)$ действительно однозначно определяются вероятности вида $P\{x(t_i) = k_i, 1 \leq i \leq n\}$, т. е. полностью характеризуется данный поток как случайный процесс.

В частности, для простейшего потока с параметром λ мы имели (§ 2)

$$v_k(t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!} \quad (k = 0, 1, 2, \dots);$$

поэтому для простейшего потока с параметром λ мы имеем при

$$0 = t_0 < t_1 < \dots < t_n, \quad 0 \leq k_0 \leq k_1 \leq \dots \leq k_n$$

$$\begin{aligned} P\{x(t_i) = k_i, 1 \leq i \leq n\} &= \\ &= \prod_{i=1}^n e^{-\lambda(t_i - t_{i-1})} \frac{\lambda^{k_i - k_{i-1}} (t_i - t_{i-1})^{k_i - k_{i-1}}}{(k_i - k_{i-1})!} = \\ &= e^{-\lambda t_n} \lambda^{k_n} \prod_{i=1}^n \frac{(t_i - t_{i-1})^{k_i - k_{i-1}}}{(k_i - k_{i-1})!}. \end{aligned}$$

§ 7. Основное свойство стационарных потоков

Для любого стационарного потока условимся, как в главе 1, обозначать через $w(t)$ вероятность того, что в течение промежутка времени длины t произойдет по крайней мере один вызов, так что

$$w(t) = 1 - v_0(t) = \sum_{k=1}^{\infty} v_k(t).$$

В § 4 мы убедились, что для стационарного потока без последствия (в частности, для простейшего потока) отношение $w(t)/t$ при $t \rightarrow 0$ стремится к определенному пределу λ , который мы называли параметром данного потока и который, как мы видели, имеет важнейшее значение для изучения основных свойств этого потока. Соотношение

$$\frac{w(t)}{t} \rightarrow \lambda \quad (t \rightarrow 0), \quad (7.1)$$

равносильное соотношению

$$w(t) = \lambda t + o(t) \quad (t \rightarrow 0), \quad (7.2)$$

часто выражают, говоря, что $w(t)$ при малых t «асимптотически пропорционально» t . В подавляющем большинстве изложений теории простейшего потока соотношение (7.2) [или (7.1)] прямо включается в определение простейшего потока*), что, как мы видели, является излишним, так как это соотношение выводится как простое следствие требований стационарности и отсутствия последействия.

Однако соотношение (7.1) на самом деле обладает еще значительно более широкой областью применимости; оно имеет место, как мы теперь убедимся, для *любого стационарного потока*. Параметром, как мы его определяли в главе 1, обладает, таким образом, каждый стационарный поток, независимо от наличия или отсутствия последействия. Это обстоятельство дает нам, как мы увидим, весьма удобный опорный пункт для изучения общих свойств стационарных потоков.

Доказательство существования предела (7.1) для любого стационарного потока опирается на следующую элементарную лемму теории пределов, которая, как мы увидим, пригодится нам и в дальнейшем.

Лемма. Пусть функция $f(x)$ — неотрицательная и неубывающая в отрезке $0 < x \leq a$ и $f(x+y) \leq f(x) + f(y)$, если x, y и $x+y$ принадлежат отрезку $(0, a)$. Тогда отношение $f(x)/x$ при $x \rightarrow 0$ либо безгранично возрастает, либо стремится к некоторому пределу; этот предел равен нулю только в тривиальном случае $f(a) = 0$.

Доказательство. Из неравенства $f(x+y) \leq f(x) + f(y)$ легко следует, что

$$f(x) \leq mf\left(\frac{x}{m}\right) \quad (7.3)$$

для $0 < x \leq a$ и любого натурального числа m ; в частности, при $x = a$

$$\frac{f\left(\frac{a}{m}\right)}{\frac{a}{m}} \geq \frac{f(a)}{a} \quad (m = 1, 2, \dots);$$

*) См. Erlang [7], Феллер [3], Фрай [4], Хинчин [5].

это показывает, что [за исключением тривиального случая $f(a)=0$]

$$\alpha = \overline{\lim}_{x \rightarrow 0} \frac{f(x)}{x} \geq \frac{f(a)}{a} > 0,$$

причем не исключен случай $\alpha = +\infty$.

Допустим сначала, что $\alpha < +\infty$. Пусть число $c > 0$ таково, что

$$\frac{f(c)}{c} > \alpha - \varepsilon, \quad (7.4)$$

где $\varepsilon > 0$ — сколь угодно малое наперед заданное число, и пусть $0 < x < c$. Определим натуральное число $m \geq 2$ из неравенств

$$\frac{c}{m} \leq x < \frac{c}{m-1};$$

тогда в силу (7.3) и предположенной монотонности $f(x)$

$$\frac{f(x)}{x} \geq \frac{f\left(\frac{c}{m}\right)}{\frac{c}{m-1}} = \frac{m-1}{m} \frac{mf\left(\frac{c}{m}\right)}{c} \geq \frac{m-1}{m} \frac{f(c)}{c}, \quad (7.5)$$

и, значит, в силу (7.4)

$$\frac{f(x)}{x} \geq \left(1 - \frac{1}{m}\right) (\alpha - \varepsilon),$$

а так как ε произвольно мало и $m \rightarrow \infty$ при $x \rightarrow 0$, то

$$\lim_{x \rightarrow 0} \frac{f(x)}{x} = \alpha,$$

и лемма доказана. Рассуждение остается в принципе тем же при $\alpha = +\infty$. Мы берем произвольно большое $A > 0$ и выбираем число c так, что $f(c)/c > A$; тогда из (7.5)

$$\frac{f(x)}{x} \geq \frac{m-1}{m} A,$$

и значит, $f(x)/x \rightarrow +\infty$ при $x \rightarrow 0$.

Теорема. Для любого стационарного потока существует

$$\lim_{t \rightarrow 0} \frac{w(t)}{t} = \lambda > 0,$$

причем не исключен случай $\lambda = +\infty$.

Доказательство. Достаточно показать, что функция $w(t)$ в некотором отрезке $(0, a)$ удовлетворяет всем предпосылкам только что доказанной леммы. Очевидно, что $w(t) \geq 0$ и с возрастанием t не может убывать; очевидно также, что в силу стационарности потока $w(a) > 0$, если мы отбросим тривиальный случай потока, в котором вызовы вообще невозможны. Наконец, если произошел по меньшей мере один вызов в промежутке $(0, t_1 + t_2)$, то, очевидно, то же самое должно иметь место хотя бы для одного из двух промежутков $(0, t_1)$ и $(t_1, t_1 + t_2)$, откуда

$$w(t_1 + t_2) \leq w(t_1) + w(t_2) \quad (t_1 > 0, t_2 > 0, t_1 + t_2 < a);$$

таким образом, и эта последняя предпосылка оказывается выполненной. Применяя лемму, мы видим, что теорема доказана.

Если данный стационарный поток есть поток без последствия, то для него, как мы это показали в § 2, мы имеем

$$v_0(t) = e^{-\lambda t} \quad (\lambda > 0 \text{ постоянная}),$$

за исключением случаев, когда в любом промежутке времени либо с достоверностью вовсе не поступает вызовов, либо с достоверностью поступает бесконечное множество вызовов; эти два случая, как не имеющие практического значения, мы условились выше исключить из рассмотрения. Отсюда для потока без последствия

$$w(t) = 1 - e^{-\lambda t}, \quad \lim_{t \rightarrow 0} \frac{w(t)}{t} = \lambda;$$

предел, существование которого мы доказали в последней теореме, в случае потока без последствия всегда есть, таким образом, некоторое конечное положительное число. Но если допустить возможность последствия, то λ может обращаться в $+\infty$ и в случаях, не исключенных нами из рассмотрения.

Чтобы в этом убедиться, рассмотрим следующий пример стационарного потока. Вообразим себе простейший поток, параметр λ которого представляет собой случайную величину, распределенную по некоторому закону $F(x)$ [$F(+0)=0$, $F(+\infty)=1$]. В конце § 6 мы определили вероятность системы равенств $x(t_i)=k_i$ ($1 \leq i \leq n$) для простейшего потока с данным значением параметра λ ; будем теперь для краткости обозначать эту вероятность через $P_\lambda(t_i, k_i)$. Если параметр λ есть случайная величина, распределенная по закону $F(x)$, то вероятность системы равенств $x(t_i)=k_i$ ($1 \leq i \leq n$) будет (по формуле полной вероятности) равна

$$P(t_i, k_i) = \int_0^\infty P_\lambda(t_i, k_i) dF(\lambda). \quad (7.6)$$

Эта вероятность определена, таким образом, для любых n, t_i, k_i ; а это, как мы знаем, означает задание определенного потока; этот поток будет, очевидно, стационарным; но, вообще говоря, он будет потоком с последствием. Если $w(t)$ и $v_k(t)$ имеют обычное значение для потока (7.6), а $w_\lambda(t)$ и $v_{k\lambda}(t)$ означают те же величины для простейшего потока с параметром λ , то мы, очевидно, имеем

$$w(t) = \int_0^\infty w_\lambda(t) dF(\lambda)$$

и

$$w_\lambda(t) = 1 - v_{0\lambda}(t) = 1 - e^{-\lambda t},$$

откуда

$$\frac{w(t)}{t} = \int_0^\infty \frac{1 - e^{-\lambda t}}{t} dF(\lambda).$$

Если

$$\int_0^\infty \lambda dF(\lambda) < +\infty$$

[т. е. закон $F(x)$ имеет конечное математическое ожидание], то в силу неравенства $1 - e^{-\lambda t} \leq \lambda t$ мы получаем

$$\frac{w(t)}{t} \leq \int_0^\infty \lambda dF(\lambda);$$

отношение $w(t)/t$ при $t \rightarrow 0$ ограничено, и рассматриваемый нами поток имеет конечный параметр. Но если интеграл

$$\int_0^{\infty} \lambda dF(\lambda) \quad (7.7)$$

расходится, то $w(t)/t \rightarrow +\infty$ при $t \rightarrow 0$. В самом деле, пусть $\varepsilon > 0$ произвольно мало и A столь велико, что

$$\int_0^A \lambda dF(\lambda) > \frac{1}{\varepsilon};$$

тогда в силу

$$\frac{w(t)}{t} \geq \int_0^A \frac{1 - e^{-\lambda t}}{t} dF(\lambda) \rightarrow \int_0^A \lambda dF(\lambda) \quad (t \rightarrow 0)$$

мы имеем

$$\lim_{t \rightarrow \infty} \frac{w(t)}{t} \geq \frac{1}{\varepsilon},$$

и значит, $w(t)/t \rightarrow +\infty$ при $t \rightarrow 0$. Таким образом, в случае расходимости интеграла (7.7) наш поток имеет бесконечное значение параметра. Вместе с тем этот поток с вероятностью 1 дает конечное число вызовов в любом конечном промежутке времени; мы имеем

$$v_k(t) = \int_0^{\infty} v_{k, \lambda}(t) dF(\lambda),$$

или, так как $v_{k, \lambda}(t) = e^{-\lambda t} (\lambda t)^k / k!$,

$$v_k(t) = \frac{t^k}{k!} \int_0^{\infty} \lambda^k e^{-\lambda t} dF(\lambda) > 0,$$

$$\sum_{k=0}^{\infty} v_k(t) = \int_0^{\infty} e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} dF(\lambda) = \int_0^{\infty} dF(\lambda) = 1.$$

§ 8. Общая форма стационарного потока без последействия

Как мы видели в главе 1, стационарный поток без последействия, если он сверх того обладает еще свойством ординарности, есть простейший поток, общая структура которого легко может быть установлена. Теперь мы поставим себе задачу найти общий вид стационарного потока без последействия, отбрасывая требование ординарности.

Как мы видели в § 6, стационарный поток без последействия однозначно определяется заданием функций $v_k(t)$ ($k=0, 1, 2, \dots$) [причем всегда $v_0(t)=e^{-\lambda t}$]. Поэтому наша задача сводится к определению общего вида семейства функций $v_k(t)$ для стационарных потоков без последействия. С этой целью мы прежде всего установим, что для любого такого потока при любом $k > 0$ отношение $v_k(t)/t$ при $t \rightarrow 0$ стремится к определенному пределу, который может быть либо нулем, либо положительным числом.

Обозначим через $\psi_k(t)$ ($k=0, 1, 2, \dots$) вероятность того, что в промежутке длины t произойдет по меньшей мере k вызовов, так что для $k=0, 1, 2, \dots$

$$\psi_k(t) = \sum_{i=k}^{\infty} v_i(t), \quad v_k(t) = \psi_k(t) - \psi_{k+1}(t),$$

$$\psi_0(t) = 1, \quad \psi_1(t) = w(t), \quad \psi_2(t) = \psi(t).$$

В § 7 мы доказали, что для любого стационарного потока отношение $\psi_1(t)/t$ при $t \rightarrow 0$ стремится к определенному пределу, конечному или бесконечному. В случае потока без последействия этот результат является тривиальным, так как $\psi_1(t) = w(t) = 1 - e^{-\lambda t}$, откуда $\psi_1(t)/t \rightarrow \lambda$ при $t \rightarrow 0$. Но зато мы теперь убедимся, что в случае стационарного потока без последействия предел отношения $\psi_k(t)/t$ при $t \rightarrow 0$ существует для любого $k > 0$. Так как $v_k(t) = \psi_k(t) - \psi_{k+1}(t)$, то отсюда будет следовать, что и предел отношения $v_k(t)/t$ при $t \rightarrow 0$ существует для любого $k > 0$. В случае простейшего потока мы имеем, конечно, $v_1(t)/t \rightarrow \lambda$, $v_k(t)/t \rightarrow 0$ ($k > 1$).

Подобно рассуждению § 7 мы начнем с доказательства одного элементарного вспомогательного предложения из теории пределов, представляющего собой некоторое усиление леммы § 7.

Лемма. Пусть функция $f(x)$ — неотрицательная и неубывающая в отрезке $0 < x \leq a$, отношение $f(x)/x$ ограничено в этом отрезке и

$$f(nx) \leq nf(x) + cn^2x^2, \quad (8.1)$$

где $c > 0$ — постоянная, n — любое натуральное число, и $0 < nx \leq a$; тогда отношение $f(x)/x$ при $x \rightarrow 0$ стремится к некоторому пределу $l \geq 0$.

Доказательство. Полагая в (8.1) $x = x_0/n$, где $0 < x_0 \leq a$, получаем

$$f(x_0) \leq nf\left(\frac{x_0}{n}\right) + cx_0^2,$$

откуда

$$\frac{f\left(\frac{x_0}{n}\right)}{\frac{x_0}{n}} \geq \frac{f(x_0)}{x_0} - cx_0. \quad (8.2)$$

Положим

$$\lim_{x \rightarrow 0} \frac{f(x)}{x} = l;$$

мы можем принять $l > 0$, так как при $l = 0$ утверждение леммы тривиально.

Пусть $\varepsilon > 0$ задано произвольно. Выберем $x_0 < \frac{\varepsilon}{c}$ и так, чтобы

$$\frac{f(x_0)}{x_0} > l - \varepsilon.$$

Пусть $0 < x < x_0$ и натуральное число $n > 1$ определяется неравенствами

$$\frac{x_0}{n} \leq x < \frac{x_0}{n-1};$$

тогда в силу (8.2)

$$\begin{aligned} \frac{f(x)}{x} &\geq \frac{f\left(\frac{x_0}{n}\right)}{\frac{x_0}{n}} = \frac{n-1}{n} \frac{f\left(\frac{x_0}{n}\right)}{\frac{x_0}{n}} \geq \frac{n-1}{n} \frac{f(x_0)}{x_0} - cx_0 \geq \\ &\geq \left(1 - \frac{1}{n}\right)(l - \varepsilon) - \varepsilon > l - 3\varepsilon, \end{aligned}$$

если x достаточно мало; так как, с другой стороны, $f(x)/x < l + \varepsilon$ для достаточно малых x , то

$$\frac{f(x)}{x} \rightarrow l \quad (x \rightarrow 0),$$

и лемма доказана.

Чтобы установить существование предела отношения $\psi_k(t)/t$ при $t \rightarrow 0$, нам надо только показать, что функция $\psi_k(t)$ при $k \geq 0$ удовлетворяет всем предпосылкам доказанной леммы. Неотрицательность и монотонность $\psi_k(t)$ в любом отрезке самоочевидны. Далее, из $\psi_k(t) \leq \psi_1(t) = w(t)$ и $w(t)/t \rightarrow \lambda \quad (t \rightarrow 0)$ вытекает ограниченность $\psi_k(t)/t$ в любом отрезке. Поэтому нам остается только убедиться, что $\psi_k(t)$ удовлетворяет соотношению (8.1).

Обозначим с этой целью через g_l верхнюю грань отношения $\psi_l(t)/t$ в области $0 < t < +\infty \quad (l \geq 1)$ и положим

$$A_k = \sum_{l=1}^{k-1} g_l g_{k-l}.$$

Мы утверждаем, что

$$\psi_k(nt) \leq n\psi_k(t) + A_k \frac{n(n-1)}{2} t^2, \quad (8.3)$$

откуда и будет следовать, что функция $\psi_k(t)$ удовлетворяет соотношению (8.1).

Неравенство (8.3) мы докажем с помощью индукции по n . При $n=1$ оно тривиально. Допустим, что для некоторого n оно имеет место. Для того чтобы в отрезке $[0, (n+1)t]$ длины $(n+1)t = nt + t$ поступило не менее k вызовов [вероятность чего равна $\psi_k\{(n+1)t\}$], необходимо, чтобы при каком-либо l ($0 \leq l \leq k$) имелось не менее l вызовов в отрезке $(0, t)$ и не менее $k-l$ вызовов в отрезке $[t, (n+1)t]$ (длины nt); поэтому

$$\begin{aligned} \psi_k[(n+1)t] &\leq \sum_{l=0}^k \psi_l(t) \psi_{k-l}(nt) = \\ &= \psi_0(t) \psi_k(nt) + \psi_k(t) \psi_0(nt) + \sum_{l=1}^{k-1} \psi_l(t) \psi_{k-l}(nt) \leq \\ &\leq \psi_k(nt) + \psi_k(t) + nt^2 \sum_{l=1}^{k-1} g_l g_{k-l} = \\ &= \psi_k(nt) + \psi_k(t) + A_k nt^2. \end{aligned}$$

В силу (8.3) отсюда

$$\psi_k[(n+1)t] \leq (n+1)\psi_k(t) + A_k \frac{n(n+1)}{2} t^2,$$

а это и есть соотношение (8.3) с $n+1$ вместо n .

Таким образом, функция $\psi_k(t)$ при $k > 0$ удовлетворяет всем предпосылкам доказанной леммы, и следовательно, при $t \rightarrow 0$ отношение $\psi_k(t)/t$, а значит, и отношение $v_k(t)/t$ стремятся к определенному пределу. Так как $w(t)/t \rightarrow \lambda > 0$ при $t \rightarrow 0$, то и отношение $v_k(t)/w(t)$ при этом имеет определенный предел.

Положим

$$\lim_{t \rightarrow 0} [v_k(t)/w(t)] = p_k \quad (k = 1, 2, \dots).$$

Отношение $v_k(t)/w(t)$ есть вероятность получить k вызовов в отрезке длины t , если известно, что в этом отрезке вызовы существуют. Предел этого отношения при $t \rightarrow 0$, т. е. число p_k , можно поэтому рассматривать как вероятность получения k вызовов в определенный момент, если известно, что в этот момент вообще вызовы происходят (такое истолкование величины p_k возможно, но, конечно, не обязательно).

Перейдем теперь к определению общего вида функций $v_k(t)$ для стационарного потока без последствия. В § 3 мы установили для такого потока [см. (3.1)] общее соотношение

$$v_k(t+\tau) = \sum_{i=0}^k v_i(\tau) v_{k-i}(t) \quad (t > 0, \tau > 0, k = 0, 1, 2, \dots);$$

так как при $\tau \rightarrow 0$

$$v_0(\tau) = e^{-\lambda\tau} = 1 - \lambda\tau + o(\tau),$$

то отсюда при $k > 0$

$$v_k(t+\tau) = (1 - \lambda\tau) v_k(t) + \sum_{i=1}^k v_i(\tau) v_{k-i}(t) + o(\tau),$$

и, следовательно,

$$\frac{v_k(t+\tau) - v_k(t)}{\tau} = -\lambda v_k(t) + \sum_{i=1}^k \frac{v_i(\tau)}{\tau} v_{k-i}(t) + o(1).$$

Но при $l > 0$ и $\tau \rightarrow 0$ по доказанному выше

$$\frac{v_l(\tau)}{\tau} = \frac{v_l(\tau)}{w(\tau)} \frac{w(\tau)}{\tau} \rightarrow \lambda p_l;$$

поэтому предельный переход доказывает существование $v'_k(t)$ и дает

$$v'_k(t) = -\lambda v_k(t) + \lambda \sum_{i=1}^k p_i v_{k-i}(t) \quad (k=1, 2, \dots). \quad (8.4)$$

Добавляя сюда очевидное соотношение

$$v'_0(t) = -\lambda v_0(t), \quad (8.5)$$

мы получаем систему уравнений, позволяющих однозначно определить систему функций $v_k(t)$. В частности, к этому ведет путь замены неизвестных функций, которым мы пользовались в § 3. Полагая, как там,

$$v_k(t) = e^{-\lambda t} u_k(t) \quad (k=0, 1, 2, \dots),$$

мы легко приводим систему (8.4) к виду

$$u'_k(t) = \lambda [p_1 u_{k-1}(t) + p_2 u_{k-2}(t) + \dots + p_k u_0(t)],$$

позволяющему рекуррентно определить все функции $u_k(t)$ [а значит, и $v_k(t)$]. Так, например, мы в силу $u_0(t) \equiv 1$ находим

$$u'_1(t) = \lambda p_1,$$

откуда

$$u_1(t) = \lambda p_1 t, \quad v_1(t) = e^{-\lambda t} \lambda p_1 t.$$

Мы не будем проводить здесь дальнейших выводов, так как значительно более простые и изящные результаты дает метод производящих функций, к применению которого мы теперь и переходим.

Положим

$$F(t, x) = \sum_{k=0}^{\infty} v_k(t) x^k;$$

отыскание системы функций $v_k(t)$ сводится к отысканию функции $F(t, x)$. Умножая соотношение (8.4) [для $k=0$ заменяемое соотношением (8.5)] на x^k и суммируя по k

от 0 до ∞ , мы находим

$$\begin{aligned}\frac{\partial F}{\partial t} &= -\lambda F + \lambda \sum_{k=1}^{\infty} x^k \sum_{l=1}^k p_l v_{k-l}(t) = \\ &= -\lambda F + \lambda \sum_{l=1}^{\infty} p_l \sum_{q=0}^{\infty} v_q(t) x^{q+l} = \\ &= -\lambda F + \lambda \sum_{l=1}^{\infty} p_l x^l \sum_{q=0}^{\infty} v_q(t) x^q,\end{aligned}$$

или, полагая

$$\sum_{l=1}^{\infty} p_l x^l = \Phi(x),$$

получаем

$$\begin{aligned}\frac{\partial F}{\partial t} &= \lambda [\Phi(x) - 1] F, \\ \frac{\partial \ln F}{\partial t} &= \lambda [\Phi(x) - 1],\end{aligned}$$

а так как при любом x

$$F(0, x) = v_0(0) = 1,$$

то интегрированием по t находим

$$F(t, x) = e^{\lambda [\Phi(x) - 1] t}, \quad (8.6)$$

и наша задача решена. Заметим, еще, что при любом t

$$F(t, 1) = \sum_{k=0}^{\infty} v_k(t) = 1,$$

вследствие чего (8.6) дает

$$\Phi(1) = \sum_{l=1}^{\infty} p_l = 1.$$

Таким образом, производящая функция $F(t, x)$ для любого стационарного потока без последствия имеет вид

(8.6), где $\lambda > 0$ и $\Phi(x) = \sum_{l=1}^{\infty} p_l x^l$, $p_l \geq 0$, $\sum_{l=1}^{\infty} p_l = 1$.

Убедимся теперь, что и обратно, если числа λ и p_l ($l=1, 2, \dots$) подчиняются только что перечисленным требованиям, то

существует стационарный поток без последействия, производящая функция которого дается формулой (8.6).

С этой целью мы допустим, что моменты времени, в которые происходят вызовы, образуют простейший поток с параметром λ , так что для вероятности $v_k^*(t)$ того, что за промежуток времени t произойдет k таких «вызывающих моментов», мы имеем обычное выражение

$$v_k^*(t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}.$$

Однако поток самих вызовов не будет, вообще говоря, простейшим, так как мы допустим, что в каждый вызывающий момент может с вероятностью, отличной от нуля, поступить и более одного вызова. Примем вероятность поступления в данный вызывающий момент ровно l вызовов равной p_l ($l=1, 2, \dots$), независимо от того, каков данный вызывающий момент и каково было течение этого потока до данного момента. Этим соглашением мы задаем некоторый определенный поток вызовов, который, очевидно, будет стационарным потоком без последействия. Покажем, что производящая функция $F(t, x)$ этого потока дается формулой (8.6).

Число вызовов, происходящих в любой вызывающий момент, мы определили как случайную величину, принимающую значение l с вероятностью p_l ($l=1, 2, \dots$); производящая функция такой величины есть

$$\sum_{l=1}^{\infty} p_l x^l = \Phi(x).$$

Возьмем теперь r любых различных между собой вызывающих моментов и обозначим через $P_r(k)$ вероятность того, что в эти r моментов в совокупности произойдет k вызовов, так что суммарное число вызовов за r вызывающих моментов есть случайная величина с производящей функцией

$$\sum_{k=0}^{\infty} P_r(k) x^k$$

[где, разумеется, $P_r(k)=0$ при $k < r$]. Но эта случайная величина есть сумма r взаимно независимых случайных величин, каждая из которых имеет производящую функцию $\Phi(x)$.

Так как при сложении взаимно независимых случайных величин их производящие функции перемножаются *), то поэтому

$$\sum_{k=0}^{\infty} P_r(k) x^k = \{\Phi(x)\}^r.$$

А так как с другой стороны, очевидно, для рассматриваемого потока

$$v_k(t) = \sum_{r=0}^{\infty} v_r^*(t) P_r(k),$$

то

$$\begin{aligned} F(t, x) &= \sum_{k=0}^{\infty} v_k(t) x^k = \sum_{k=0}^{\infty} x^k \sum_{r=0}^{\infty} v_r^*(t) P_r(k) = \\ &= \sum_{r=0}^{\infty} v_r^*(t) \sum_{k=0}^{\infty} P_r(k) x^k = \sum_{r=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^r}{r!} \{\Phi(x)\}^r = \\ &= e^{-\lambda t} \sum_{r=0}^{\infty} \frac{\{\lambda \Phi(x)\}^r}{r!} = e^{\lambda t [\Phi(x) - 1]}, \end{aligned}$$

что совпадает с формулой (8.6).

Мы можем формулировать результат нашего исследования так: совокупность всех стационарных потоков без последствия совпадает с совокупностью всех потоков, даваемых формулой (8.6), где $\lambda > 0$:

$$\Phi(x) = \sum_{l=1}^{\infty} p_l x^l, \quad p_l \geq 0 \quad (l=1, 2, \dots), \quad \sum_{l=1}^{\infty} p_l = 1.$$

С предметной точки зрения мы убедились, что для каждого стационарного потока без последствия поток вызывающих моментов является простейшим и для полного описания данного потока вызовов надо, кроме параметра λ этого простейшего потока, задать еще закон распределения $(p_1, p_2, \dots, p_n, \dots)$ числа вызовов, поступающих в любой

*) Это непосредственно вытекает из того, что для случайной величины ξ с законом распределения $P\{\xi=n\} = q_n$ ($n=0, 1, 2, \dots$) производящая функция $f(x) = \sum_{n=0}^{\infty} q_n x^n$ есть, очевидно, математическое ожидание величины x^ξ .

выбранный вызывающий момент. Очевидно, эти соображения делают совершенно прозрачной структуру самого общего стационарного потока без последействия.

Заметим еще, что в случае $p_1 = 1$, $p_k = 0$ ($k > 1$) формула (8.6) дает нам производящую функцию простейшего потока с параметром λ :

$$F(t, x) = e^{\lambda t(x-1)} = \sum_{k=0}^{\infty} \left\{ e^{-\lambda t} \frac{(\lambda t)^k}{k!} \right\} x^k.$$

Глава 3

ФУНКЦИИ ПАЛЬМА

§ 9. Определение и доказательство существования

После того как в конце предыдущей главы мы полностью выяснили строение стационарных потоков без последействия, мы должны теперь обратиться к исследованию потоков более общего типа. Для довольно широких классов таких потоков весьма удобным орудием исследования оказалась одна функция, введенная Пальмом [8] и примененная им с успехом к решению ряда задач. Пальм определяет эту функцию $\varphi_0(t)$ (для любого стационарного потока) как *условную вероятность отсутствия вызовов в промежутке $(t_0, t_0 + t)$, если известно, что в момент t_0 произошел вызов*. Однако такое определение вряд ли можно считать достаточно удобным; то условие, при котором должна быть вычислена вероятность $\varphi_0(t)$, т. е. наличие вызова в некоторый момент t_0 , само имеет во всех актуальных случаях вероятность 0, и это обстоятельство, как известно, не позволяет нам непосредственно определить функцию $\varphi_0(t)$ для заданного потока с помощью известных правил расчета условных вероятностей. Поэтому мы дадим этой функции другое, более сложное определение, которое позволит однозначно определять ее для любого стационарного потока. Вместе с тем мы определим не одну функцию, а целую последовательность $\varphi_k(t)$ ($k = 0, 1, 2, \dots$) функций, которые будем называть функциями Пальма и которые в дальнейшем окажутся нам полезными при решении ряда важных задач.

Пусть мы имеем два последовательных промежутка времени, из которых первый имеет длину τ , а второй — t (в дальнейшем мы будем для краткости называть самые эти промежутки соответственно «промежутком τ » и «промежутком t »). Обозначим для данного стационарного потока через $H_k(\tau, t)$ ($k \geq 0$) вероятность следующего двойного события: 1) в промежутке τ произойдет по меньшей мере один вызов; 2) в промежутке t произойдет не более k вызовов. Эти два события, вообще говоря, будут взаимно зависимы; так как вероятность события 1) в наших старых обозначениях есть $w(\tau)$, то отношение

$$\frac{H_k(\tau, t)}{w(\tau)} \quad (9.1)$$

выражает собою условную вероятность события 2) при условии, что имело место событие 1), т. е. вероятность появления не более k вызовов в промежутке t при условии, что в промежутке τ появился по меньшей мере один вызов.

Если это отношение при $\tau \rightarrow 0$ (и при постоянном t) стремится к некоторому пределу, то этот предел естественно называть условной вероятностью появления не более чем k вызовов в промежутке t при условии, что в начальный момент этого промежутка произошел вызов.

Убедимся теперь, что предел отношения (9.1) при $\tau \rightarrow 0$ (и постоянном t) всегда существует, если только данный стационарный поток имеет конечный параметр λ . С этой целью рассмотрим сначала отношение $H_k(\tau, t)/t$. Чтобы доказать существование предела этого отношения при $\tau \rightarrow 0$, достаточно убедиться, что величина $H_k(\tau, t)$ как функция от τ удовлетворяет всем предпосылкам леммы § 7. Неотрицательность и монотонность этой функции самоочевидны. Пусть $\tau = \tau_1 + \tau_2$ и промежуток τ_1 предшествует промежутку τ_2 . Тогда, если выполнено то двойное событие, вероятность которого мы обозначили $H_k(\tau, t)$, то, очевидно, выполняется по меньшей мере одно из следующих двух событий:

(А) В промежутке τ_2 имеется по меньшей мере один вызов, в промежутке t имеется не более k вызовов [вероятность события (А) равна $H_k(\tau_2, t)$].

(В) В промежутке τ_1 имеется по меньшей мере один вызов, в промежутке $\tau_2 + t$ имеется не более k вызовов

[вероятность события (В) равна $H_k(\tau_1, \tau_2 + t) \leq H_k(\tau_1, t)$ (так как при фиксированном τ , $H_k(\tau, t)$, очевидно, есть невозрастающая функция от t)]. Таким образом, мы находим

$$H_k(\tau_1 + \tau_2, t) \leq H_k(\tau_1, t) + H_k(\tau_2, t),$$

т. е. функция $H_k(\tau, t)$ удовлетворяет (относительно τ) и последней предпосылке леммы § 7. Применяя эту лемму, мы находим, что отношение $H_k(\tau, t)/\tau$ при $\tau \rightarrow 0$ стремится к некоторому пределу или безгранично возрастает; однако последний случай исключается, так как, очевидно, всегда $H_k(\tau, t) \leq w(\tau)$, а отношение $w(\tau)/\tau$ по нашему предположению стремится к конечному пределу λ .

Наконец,

$$\frac{H_k(\tau, t)}{w(\tau)} = \frac{H_k(\tau, t)/\tau}{w(\tau)/\tau};$$

числитель и знаменатель этой дроби по доказанному стремятся при $\tau \rightarrow 0$ к определенным пределам; поэтому и

$$\lim_{\tau \rightarrow 0} \frac{H_k(\tau, t)}{w(\tau)} = \Phi_k(t) \quad (9.2)$$

существует; разумеется, этот предел является функцией от t .

Положим теперь

$$h_0(\tau, t) = H_0(\tau, t), \quad h_k(\tau, t) = H_k(\tau, t) - H_{k-1}(\tau, t) \quad (k > 0);$$

очевидно, $h_k(\tau, t)$ есть вероятность того, что 1) в промежутке τ имеется по меньшей мере один вызов и 2) в промежутке t имеется ровно k вызовов; отношение $h_k(\tau, t)/w(\tau)$ представляет собой условную вероятность иметь k вызовов в промежутке t при условии, что в промежутке τ имеется по меньшей мере один вызов. Из (9.2) следует

$$\lim_{\tau \rightarrow 0} \frac{h_0(\tau, t)}{w(\tau)} = \Phi_0(t), \quad \lim_{\tau \rightarrow 0} \frac{h_k(\tau, t)}{w(\tau)} = \Phi_k(t) - \Phi_{k-1}(t) \quad (k > 0).$$

Полагая

$$\varphi_0(t) = \Phi_0(t), \quad \varphi_k(t) = \Phi_k(t) - \Phi_{k-1}(t) \quad (k > 0).$$

мы имеем

$$\lim_{\tau \rightarrow 0} \frac{h_k(\tau, t)}{w(\tau)} = \varphi_k(t) \quad (k = 0, 1, 2, \dots).$$

Функции $\Phi_k(t)$ мы и будем называть функциями Пальма. Функция $\Phi_k(t)$ может быть понимаема как *вероятность иметь k вызовов в промежутке длины t при условии, что в начальный момент этого промежутка произошел вызов*. Этим она отличается от функции $v_k(t)$, представляющей собою вероятность того же события при условии, что относительно начального момента ничего неизвестно. Проведенное нами рассуждение показывает, что вся совокупность функций Пальма однозначно определяется для любого стационарного потока с конечным параметром λ .

Заметим еще, что так как $H_k(\tau, t)$ относительно t есть функция невозрастающая, а $w(\tau)$ от t не зависит, то все функции $\Phi_k(t)$ [в частности, функция $\Phi_0(t) = \varphi_0(t)$] — невозрастающие в области $0 < t < +\infty$.

§ 10. Формулы Пальма

Функции Пальма связаны с основными функциями $v_k(t)$ данного стационарного потока простыми и важными формулами, которые нам предстоит теперь вывести.

Допустим, что данный стационарный поток — ординарный и имеет конечный параметр λ . Рассмотрим снова промежуток времени длины $\tau + t$, составленный из «промежутка τ » и непосредственно следующего за ним «промежутка t ». Обозначим через n_1 и n_2 соответственно числа вызовов в промежутках τ и t (n_1, n_2 — случайные величины). Мы имеем, очевидно,

$$v_k(\tau + t) = P\{n_1 + n_2 = k\} = \sum_{r=0}^k P\{n_1 = r, n_2 = k - r\},$$

откуда в силу ординарности данного потока при $\tau \rightarrow 0$

$$v_k(\tau + t) = P\{n_1 = 0, n_2 = k\} + P\{n_1 = 1, n_2 = k - 1\} + o(\tau). \quad (10.1)$$

$$P\{n_1 = 0, n_2 = k\} = P\{n_2 = k\} - P\{n_1 > 0, n_2 = k\} = v_k(t) - h_k(\tau, t), \quad (10.2)$$

где мы пользуемся обозначениями § 9. С другой стороны,

$$P\{n_1 = 1, n_2 = k - 1\} = P\{n_1 > 0, n_2 = k - 1\} - P\{n_1 > 1, n_2 = k - 1\} = h_{k-1}(\tau, t) + o(\tau). \quad (10.3)$$

Вставляя (10.2) и (10.3) в (10.1), находим

$$v_k(t + \tau) = v_k(t) - h_k(\tau, t) + h_{k-1}(\tau, t) + o(\tau),$$

откуда

$$\frac{v_k(\tau + t) - v_k(t)}{\tau} = \frac{h_{k-1}(\tau, t)}{w(\tau)} \frac{w(\tau)}{\tau} - \frac{h_k(\tau, t)}{w(\tau)} \frac{w(\tau)}{\tau} + o(1).$$

В силу результатов § 9 отсюда вытекает дифференцируемость функции $v_k(t)$ и соотношение

$$v'_k(t) = \lambda [\varphi_{k-1}(t) - \varphi_k(t)] \quad (k > 0); \quad (10.4)$$

при $k=0$ это же рассуждение дает

$$v'_0(t) = -\lambda \varphi_0(t), \quad (10.5)$$

так что соотношение (10.4) имеет место и при $k=0$, если положить $\varphi_{-1}(t) \equiv 0$. Складывая соотношения (10.4) для $k=0, 1, \dots, m$ и обозначая через

$$V_m(t) = \sum_{k=0}^m v_k(t)$$

вероятность иметь в промежутке длины t не более m вызовов, мы находим

$$V'_m(t) = -\lambda \varphi_m(t) \quad (m=0, 1, 2, \dots). \quad (10.6)$$

Формулы (10.4) и (10.6) и были целью нашего вывода [у Пальма имеется лишь формула (10.5)]. Эти формулы иногда удобнее применять в интегральной форме. Интегрируя обе части (10.6) от 0 до t мы находим

$$V_m(+0) - V_m(t) = \lambda \int_0^t \varphi_m(u) du \quad (m=0, 1, 2, \dots),$$

но

$$V_m(+0) = 1;$$

в самом деле, мы имеем

$$V_m(+0) \geq V_0(+0) = v_0(+0) = 1 - w(+0);$$

а так как, при $t \rightarrow 0$, $w(t)/t \rightarrow \lambda$, то $w(+0) = 0$. Итак, мы находим для любого $t > 0$

$$1 - V_m(t) = \lambda \int_0^t \varphi_m(u) du \quad (m = 0, 1, 2, \dots), \quad (10.7)$$

откуда легко вытекает

$$\left. \begin{aligned} v_0(t) &= 1 - \lambda \int_0^t \varphi_0(u) du; \\ v_k(t) &= \lambda \int_0^t [\varphi_{k-1}(u) - \varphi_k(u)] du \quad (k = 1, 2, \dots). \end{aligned} \right\} \quad (10.8)$$

Эти формулы просто и непосредственно выражают функции $v_k(t)$ данного потока через функции Пальма $\varphi_k(t)$.

§ 11. Интенсивность стационарного потока. Теорема Королюка

В § 4 мы условились называть интенсивностью μ данного стационарного потока математическое ожидание числа вызовов в единицу времени; в силу аддитивности математических ожиданий мы имеем тогда, что математическое ожидание числа вызовов в промежутке длины t для стационарного потока пропорционально t , т. е.,

$$\sum_{k=0}^{\infty} k v_k(t) = \mu t.$$

Там же мы убедились, что всегда $\mu \geq \lambda$, а для простейшего потока $\mu = \lambda$.

В работах прикладного характера совпадение параметров μ и λ обычно принимается как самоочевидный факт, не требующий даже оговорок, при исследовании стационарных потоков самого общего типа. Ввиду практического значения этого допущения представляется важным разобраться в его предпосылках и дать ему строгое обоснование там, где это возможно.

Остановимся сначала на случае стационарного потока без последствия. В § 8 мы видели, что для потоков этого рода

производящая функция

$$F(t, x) = \sum_{k=0}^{\infty} v_k(t) x^k$$

имеет вид

$$e^{\lambda t [\Phi(x) - 1]},$$

где $\lambda > 0$, $\Phi(x) = \sum_{l=1}^{\infty} p_l x^l$, $p_l \geq 0$ ($l = 1, 2, \dots$), $\sum_{l=1}^{\infty} p_l = 1$.

Так как, очевидно,

$$\mu t = \sum_{k=0}^{\infty} k v_k(t) = \left\{ \frac{\partial F(t, x)}{\partial x} \right\}_{x=1}$$

и так как

$$\frac{\partial F(t, x)}{\partial x} = F(t, x) \lambda t \Phi'(x), \quad F(t, 1) = \sum_{k=0}^{\infty} v_k(t) = 1,$$

то

$$\mu t = \lambda t \Phi'(1) = \lambda t [p_1 + 2p_2 + 3p_3 + \dots],$$

откуда

$$\mu = \lambda [p_1 + 2p_2 + 3p_3 + \dots].$$

Так как

$$\sum_{l=1}^{\infty} p_l = 1$$

и

$$p_l \geq 0 \quad (l \geq 1),$$

то мы непосредственно видим, что для равенства $\mu = \lambda$ необходимо и достаточно иметь $p_1 = 1$. Но при $p_1 = 1$ данный поток, как мы видели в конце § 8, является простейшим. Таким образом, среди стационарных потоков без последействия только простейшие потоки удовлетворяют требованию $\mu = \lambda$; для всех остальных $\mu > \lambda$.

Так как для стационарного потока без последействия ординарность есть необходимое и достаточное условие для того, чтобы этот поток был простейшим, то можно еще сказать, что для стационарного потока без последействия, интенсивность которого конечна, необходимым и достаточным

условием равенства $\mu = \lambda$ является ординарность этого потока *).

Выведенные нами в § 10 формулы Пальма позволяют, как это показал В. С. Королюк, легко убедиться, что для любого стационарного потока ординарность влечет за собой равенство $\mu = \lambda$ (причем не исключается случай $\mu = \lambda = +\infty$).

В самом деле, мы имеем

$$\begin{aligned}\mu &= \sum_{k=1}^{\infty} k v_k(1) = \sum_{k=1}^{\infty} [v_k(1) + v_{k+1}(1) + v_{k+2}(1) + \dots] = \\ &= \sum_{k=1}^{\infty} [1 - v_{k-1}(1)] = \sum_{k=0}^{\infty} [1 - v_k(1)],\end{aligned}$$

откуда в силу формулы (10.7)

$$\mu = \lambda \sum_{k=0}^{\infty} \int_0^1 \varphi_k(u) du. \quad (11.1)$$

Но

$$\varphi_k(u) = \lim_{\tau \rightarrow 0} \frac{h_k(\tau, u)}{w(\tau)}, \quad (11.2)$$

а так как отношение $h_k(\tau, u)/w(\tau)$ есть условная вероятность иметь в промежутке u ровно k вызовов (при условии наличия вызовов в промежутке τ), то при любом $m > 0$

$$\sum_{k=0}^m \frac{h_k(\tau, u)}{w(\tau)} \leq 1 \quad (0 < u \leq 1),$$

а потому в силу (11.2)

$$\sum_{k=0}^m \varphi_k(u) \leq 1 \quad (0 < u \leq 1);$$

*) Совсем простое доказательство этого предложения и пример, в силу которого для потоков с бесконечной интенсивностью из равенства $\mu = \lambda$ ординарность потока не вытекает, были даны Ф. Зитеком в работе «Заметка об одной теореме Королюка», Чехословац. матем. журнал, т. 7 (82), 1957, 318—319.— Б. Г.

следовательно, при любом $m > 0$

$$\sum_{k=0}^m \int_0^1 \varphi_k(u) du = \int_0^1 \left\{ \sum_{k=0}^m \varphi_k(u) \right\} du \leq 1;$$

а значит, и

$$\sum_{k=0}^{\infty} \int_0^1 \varphi_k(u) du \leq 1,$$

и (11.1) дает $\mu \leq \lambda$; а так как еще в § 4 мы видели, что всегда $\mu \geq \lambda$, то $\mu = \lambda$, и наше утверждение доказано.

Глава 4

ПОТОКИ С ОГРАНИЧЕННЫМ ПОСЛЕДЕЙСТВИЕМ

§ 12. Другой способ описания потока

Способ задания потока вызовов, описанный нами в § 6, исходит из понимания потока как случайного процесса $x(t)$ и ничем не отличается от общего способа описания произвольного случайного процесса. Это автоматическое включение теории потоков в общую теорию случайных процессов, несомненно, имеет свои преимущества, так как позволяет применять к изучению потоков методы и результаты общей теории случайных процессов; однако, учитывая специфические свойства наших потоков как случайных процессов [и в первую очередь то, что величина $x(t)$ всегда монотонна и принимает лишь целые неотрицательные значения], мы можем надеяться найти для этих потоков хоть и менее общий, но зато более простой и удобный способ описания. Этим вопросом мы теперь и займемся.

Пусть снова начальный момент данного потока есть $t_0 = 0$; пусть t_i ($i = 1, 2, \dots$) есть момент i -го вызова, так что $t_{i-1} \leq t_i$ ($i = 1, 2, \dots$); положим, наконец,

$$t_i - t_{i-1} = z_i \quad (i = 1, 2, \dots),$$

так что $z_1 = t_1$, а z_i при $i > 1$ означает величину промежутка времени между $(i-1)$ -м и i -м вызовами. Очевидно, все t_i и z_i ($i = 1, 2, \dots$) представляют собой случайные величины, способные принимать лишь неотрицательные значения.

Условимся теперь считать поток заданным, если для любого $n > 0$ задан n -мерный закон распределения вектора $(z_1, z_2, \dots, z_n)$. Очевидно, этот способ описания потока более элементарен, чем выбранный нами в § 6, так как в основе его лежит представление о потоке не как о случайном процессе общего вида, а как о последовательности случайных величин. Убедимся теперь, что оба способа задания потока равносильны, т. е. что поток, заданный с помощью какого-либо одного из этих двух способов, будет тем самым однозначно описанным и в смысле другого способа.

Пусть, как в § 6, $x(t)$ означает число вызовов, предшествующих моменту t ; очевидно, неравенства $t_k < u$ и $x(u) \geq k$ выражают собой одно и то же событие; то же самое имеет место и для неравенств $t_k \geq u$, $x(u) < k$, а значит, и для неравенств $u \leq t_k < v$, $x(u) < k \leq x(v)$. Отсюда далее следует, что система неравенств

$$u_k \leq t_k < v_k \quad (1 \leq k \leq n) \quad (12.1)$$

выражает то же событие, что и система неравенств

$$x(u_k) < k \leq x(v_k) \quad (1 \leq k \leq n), \quad (12.2)$$

каковы бы ни были вещественные числа u_k, v_k ($1 \leq k \leq n$).

Если поток вызовов задан в смысле § 6, то при любом n и при любых u_k, v_k ($1 \leq k \leq n$) нам задан закон распределения $2n$ -мерного вектора $[x(u_k), x(v_k); k=1, 2, \dots, n]$ и, значит, однозначно определена вероятность системы (12.2), а следовательно, и равносильной ей системы (12.1). Но ввиду произвольности чисел u_k, v_k последнее означает, что однозначно задан закон распределения вектора $(t_1, t_2, \dots, t_n)$, а значит, и вектора

$$(z_1, z_2, \dots, z_n) = (t_1, t_2 - t_1, \dots, t_n - t_{n-1});$$

таким образом, данный поток однозначно определен и в новом смысле.

Обратно, если при любом n нам задан закон распределения вектора $(z_1, z_2, \dots, z_n)$, то в силу

$$t_k = \sum_{i=1}^k z_i \quad (1 \leq k \leq n)$$

тем самым однозначно определен и закон распределения вектора $(t_1, t_2, \dots, t_n)$. Но при любых $u_1, u_2, \dots, u_n$ и любых

целых $k_1, k_2, \dots, k_n$ система неравенств

$$t_{k_i} < u_i, \quad t_{k_{i+1}} \geq u_i \quad (1 \leq i \leq n) \quad (12.3)$$

равносильна системе равенств

$$x(u_i) = k_i \quad (1 \leq i \leq n); \quad (12.4)$$

так как вероятность системы (12.3) однозначно определена, то то же имеет место и для системы (12.4). А это означает, что при любом n и любых u_i ($1 \leq i \leq n$) однозначно определен закон распределения вектора $[x(u_1), x(u_2), \dots, x(u_n)]$, т. е. что наш процесс однозначно определен в смысле § 6.

Таким образом, указанный нами новый способ задания потока вызовов действительно равносильен принятому в § 6.

§ 13. Потоки с ограниченным последствием

Если данный поток — без последствия, то величины $z_1, z_2, \dots, z_n$, очевидно, взаимно независимы. Однако обратное заключение, как мы узнаем в дальнейшем, было бы неверным. Взаимная независимость величин z_k в значительной степени ограничивает явление последствия, но не исключает его полностью. В части II мы узнаем, что как раз потоки с взаимно независимыми z_k , но с наличием последствия играют важнейшую роль в теории обслуживания вызовов полнодоступным пучком линий. Мы должны поэтому заняться теперь установлением некоторых основных свойств таких потоков.

Условимся (следуя Пальму) называть потоком с ограничением последствием всякий поток, у которого $z_1, z_2, \dots, z_n, \dots$ есть последовательность взаимно независимых случайных величин*). Очевидно, для однозначного описания такого потока достаточно задать законы распределения всех величин z_k ($k = 1, 2, \dots$). Мы будем в дальнейшем обозначать эти законы через $F_k(x)$ ($k = 1, 2, \dots$).

Мы знаем, что если данный поток — стационарный и ординарный, то полное отсутствие последствия влечет за собой простейший характер потока, подробно изученный нами

*) В другом плане потоки с ограниченным последствием изучаются в теории восстановления, см. работу В. Смита «Теория восстановления и смежные с ней вопросы», сб. переводов «Математика», т. 5, вып. 3, 1961, 95—150.— Б. Г.

в главе 1. Поэтому стационарный и ординарный поток с ограниченным последствием мы можем рассматривать как некоторое обобщение простейшего потока *). Именно такого рода потоки представляют значительный интерес для теории обслуживания в случае систем с потерями (см. далее гл. 8). Стационарный ординарный поток с ограниченным последствием мы будем для краткости называть потоком типа P (или потоком Пальма).

В § 9 мы ввели для любого стационарного потока систему «функций Пальма» $\Phi_k(t)$ ($k=0, 1, 2, \dots$). Функция $\Phi_0(t)$, как мы теперь увидим, играет основную роль в теории потоков типа P . *Заданием этой функции поток типа P однозначно определяется.* В самом деле, так как в случае потока типа P величины $z_1, z_2, \dots, z_n, \dots$ между собою независимы, то для однозначного определения закона распределения каждого вектора $(z_1, z_2, \dots, z_n)$ ($n=1, 2, \dots$), а значит, и для однозначного описания потока достаточно задать законы распределения $F_k(x)$ величин z_k ($k=1, 2, \dots$). Но эти законы однозначно определяются заданием функции Пальма $\Phi_0(t)$, как показывает следующее предложение.

Теорема. Для потока типа P

$$F_1(x) = \lambda \int_0^x \Phi_0(u) du, \quad F_k(x) = 1 - \Phi_0(x) \quad (k \geq 2).$$

Для лучшей обозримости мы разобьем доказательство на несколько этапов.

1. Так как $F_1(x) = P\{t_1 < x\}$ есть вероятность наличия вызовов в промежутке $(0, x)$ и, следовательно, в наших старых обозначениях равна

$$w(x) = 1 - v_0(x) = 1 - V_0(x),$$

то утверждаемое выражение для $F_1(x)$ непосредственно вытекает из формулы (10.7) при $m=0$; при этом параметр λ

*) Интересно заметить, что ограниченность последствия следует из отсутствия последствия лишь для ординарных потоков. Неординарный поток без последствия может не быть потоком с ограниченным последствием. Примером может служить поток, рассмотренный нами в конце § 8 при $\rho_1 = \rho_2 = \frac{1}{2}$. (Этим замечанием автор обязан П. И. Васильеву.)

данного потока определяется через функцию $\Phi_0(t)$ с помощью соотношения

$$\lambda \int_0^{\infty} \Phi_0(u) du = F_1(+\infty) = 1.$$

Нам остается, таким образом, рассмотреть случай $k > 1$.

2. Положим, как в § 8,

$$\Psi_k(t) = 1 - V_{k-1}(t) = v_k(t) + v_{k+1}(t) + \dots$$

Тогда имеет место

Лемма. (См. примечание на стр. 68 — Б. Г.) Для любого потока типа P и любого $r > 0$

$$\frac{\Psi_{r+1}(u)}{\Psi_r(u)} \rightarrow 0 \quad (u \rightarrow 0).$$

Доказательство. Обозначая через t_k момент k -го вызова и полагая, как прежде, $t_k - t_{k-1} = z_k$ ($k = 1, 2, \dots$), мы очевидно имеем

$$\Psi_{r+1}(u) = P\{t_{r+1} < u\} \leq P\{t_r < u, z_{r+1} < u\},$$

и следовательно, в силу независимости z_{r+1} от t_r ,

$$\Psi_{r+1}(u) \leq \Psi_r(u) F_{r+1}(u).$$

Для доказательства леммы достаточно поэтому убедиться что $F_{r+1}(u) \rightarrow 0$ при $u \rightarrow 0$.

Пусть $a > 0$ столь велико, что $\Psi_r(a) > 0$. Пусть $x > 0$ произвольно мало и n таково, что $(n-1)x < a \leq nx$. Условимся называть «ячейками» отрезки $[(k-1)x, kx]$ ($1 \leq k \leq n$). Если $t_r < nx$ и $z_{r+1} < x$, то моменты t_r и t_{r+1} лежат либо в одной ячейке, либо в двух соседних ячейках, так что по меньшей мере один из n отрезков

$$[(l-1)x, (l+1)x] \quad (l = 1, 2, \dots, n)$$

длины $2x$ содержит более одного вызова. Поэтому

$$\begin{aligned} P\{t_r < nx, z_{r+1} < x\} &= \Psi_r(nx) F_{r+1}(x) \leq n\Psi_r(2x) = \\ &= 2nx \frac{\Psi_r(2x)}{2x} < 2(a+x) \frac{\Psi_r(2x)}{2x}, \end{aligned}$$

откуда при $x < a$

$$F_{r+1}(x) \leq \frac{2(a+x)}{\Psi_r(nx)} \frac{\Psi_r(2x)}{2x} \leq \frac{4a}{\Psi_r(a)} \frac{\Psi_r(2x)}{2x} \rightarrow 0$$

при $x \rightarrow 0$ в силу ординарности данного потока. Этим завершено доказательство нашей леммы.

3. Переходя теперь к доказательству теоремы, убедимся прежде всего, что $F_2(t) = 1 - \Phi_0(t)$. С этой целью рассмотрим введенную нами в главе 3 вероятность $h_0(\tau, t)$ того, что в некотором промежутке длины τ вызовы имеются, а в последующем за ним промежутке длины t вызовов нет. Если число вызовов в промежутке τ равно k , то из отсутствия вызовов в промежутке t следует $z_{k+1} > t$, вероятность чего есть $1 - F_{k+1}(t)$; поэтому

$$h_0(\tau, t) \leq \sum_{k=1}^{\infty} v_k(\tau) [1 - F_{k+1}(t)] \leq v_1(\tau) [1 - F_2(t)] + \psi_2(\tau).$$

С другой стороны, при том же условии (k вызовов в промежутке τ) из $z_{k+1} > t + \tau$ следует, что в промежутке t вызовов нет, поэтому

$$h_0(\tau, t) \geq \sum_{k=1}^{\infty} v_k(\tau) [1 - F_{k+1}(t + \tau)] \geq v_1(\tau) [1 - F_2(t + \tau)].$$

Таким образом,

$$v_1(\tau) [1 - F_2(t + \tau)] \leq h_0(\tau, t) \leq v_1(\tau) [1 - F_2(t)] + \psi_2(\tau).$$

Деля все части этих неравенств на $w(\tau)$ и замечая, что при $\tau \rightarrow 0$

$$\frac{v_1(\tau)}{w(\tau)} \rightarrow 1, \quad \frac{\psi_2(\tau)}{w(\tau)} \rightarrow 0, \quad \frac{h_0(\tau, t)}{w(\tau)} \rightarrow \Phi_0(t),$$

мы в пределе находим

$$1 - F_2(t + 0) \leq \Phi_0(t) \leq 1 - F_2(t),$$

откуда $F_2(t) = 1 - \Phi_0(t)$ во всех точках непрерывности закона распределения $F_2(t)$.

4. Теперь мы с помощью индукции убедимся, что $F_{r+2}(t) = 1 - \Phi_0(t)$ для всех $r \geq 0$.

В силу стационарности данного потока закон распределения расстояния между двумя первыми вызовами, следующими за каким-либо моментом $a > 0$, совпадает с законом распределения $F_2(t)$ расстояния z_2 между первыми двумя вызовами, следующими за моментом 0. Но если в промежутке $(0, a)$ имеется r вызовов, то расстояние между первыми двумя вы.

зовами, следующими за моментом a , есть z_{r+2} , и закон распределения его, равный $F_{r+2}(t)$, не зависит от предшествующего течения потока. Таким образом,

$$F_2(t) = \sum_{r=0}^{\infty} v_r(a) F_{r+2}(t). \quad (13.1)$$

Пусть теперь уже установлено, что

$$F_2(t) = F_3(t) = \dots = F_{r+1}(t) = 1 - \varphi_0(t).$$

Тогда (13.1) дает

$$1 - \varphi_0(t) = [1 - \varphi_0(t)] \sum_{k=0}^{r-1} v_k(a) + v_r(a) F_{r+2}(t) + \\ + \sum_{k>r} v_k(a) F_{k+2}(t),$$

откуда в силу $v_r(a) = \psi_r(a) - \psi_{r+1}(a)$

$$[1 - \varphi_0(t)] \psi_r(a) - F_{r+2}(t) \psi_r(a) = \\ = -\psi_{r+1}(a) F_{r+2}(t) + \sum_{k>r} v_k(a) F_{k+2}(t).$$

Так как последняя сумма правой части не превосходит

$$\sum_{k>r} v_k(a) = \psi_{r+1}(a),$$

то мы получаем

$$\psi_r(a) |1 - \varphi_0(t) - F_{r+2}(t)| \leq \psi_{r+1}(a);$$

$$|1 - \varphi_0(t) - F_{r+2}(t)| \leq \frac{\psi_{r+1}(a)}{\psi_r(a)}.$$

Это неравенство имеет место при любых $t > 0$ и $a > 0$; но при $a \rightarrow 0$ правая часть по доказанной лемме стремится к нулю; а так как левая часть от a не зависит, то при любом $t > 0$

$$F_{r+2}(t) = 1 - \varphi_0(t),$$

что и требовалось доказать.

Глава 5

ПРЕДЕЛЬНАЯ ТЕОРЕМА

§ 14. Постановка задачи. Теорема Пальма

Как мы уже говорили в главе 1, значительное большинство исследований прикладного характера основывается на предположении, что первичный поток поступающих на данную установку вызовов является простейшим потоком. Однако давно уже известен целый ряд принципиальных соображений, заставляющих сомневаться в том, что предпосылки, которые представляют собой определение простейшего потока, с достаточной степенью точности выполняются в большинстве практически встречающихся случаев (в особенности это относится к требованию отсутствия последействия). Если поэтому наблюдения и опыт констатируют некоторое небольшое отклонение реально встречающихся потоков от простейших, то этому не следует удивляться; более того, удивление может вызвать тот факт, что отклонения такого рода в большинстве случаев бывают менее значительными, чем этого можно было бы ожидать из теоретических соображений. Таким образом, если обычно при сопоставлении выводов теории с опытными данными перед исследователем встает задача — объяснить причины отклонения реально протекающих явлений от теоретически предсказанного их течения, то в данном случае дело обстоит как раз наоборот: опытные данные согласуются с выводами построенной теории, как правило, лучше, чем этого можно было бы ожидать по принципиальным соображениям, и именно это «слишком хорошее» согласие требует объяснения.

Пальмом [8] сделана заслуживающая внимания попытка объяснения фактов этого рода, исходя из предположения, что данный поток представляет собой простую сумму (суперпозицию) большого числа взаимно независимых потоков малой интенсивности, причем каждый из слагаемых потоков является стационарным и ординарным, в отношении же последействия эти потоки могут вести себя произвольным образом. При этом оказывается, что в весьма широких предположениях суммарный поток по своему характеру должен быть близок к простейшему. Такая постановка задачи, по-видимому, во многих случаях близка к реальной ситуации. Так, есл

к данной установке прикреплено большое число абонентов, то общий поток вызовов складывается из потоков (сравнительно весьма малой интенсивности), исходящих от отдельных абонентов, причем эти складываемые потоки можно в первом приближении считать стационарными, ординарными и взаимно независимыми.

Мы приходим на этом пути к ряду своеобразных предельных теорем, которые способны в значительной степени объяснить исследуемое явление. Этим вопросом мы и займемся в настоящей главе.

Пусть исследуемый поток представляет собой суперпозицию n стационарных, ординарных и взаимно независимых потоков. Обозначим через λ_r интенсивность r -го потока, через $\varphi_r(t)$ его функцию Пальма [которую в гл. 3 мы обозначали через $\varphi_0(t)$] и через $v_{kr}(t)$ — вероятность поступления в промежутке $(0, t)$ k вызовов r -го потока. Те же величины для суммарного потока обозначим соответственно через Λ , $\Phi(t)$ и $V_k(t)$ (так что в частности $\Lambda = \lambda_1 + \lambda_2 + \dots + \lambda_n$). Мы будем исходить из следующих предпосылок:

- 1°. При $n \rightarrow \infty$, Λ остается постоянным, в то время как числа $\lambda_1, \lambda_2, \dots, \lambda_n$ равномерно стремятся к нулю, так что для любого $\varepsilon > 0$ мы имеем $\lambda_r < \varepsilon$ ($r = 1, 2, \dots, n$), если n достаточно велико.
- 2°. При любом постоянном $t > 0$ и при $n \rightarrow \infty$ числа $\varphi_r(t)$ ($r = 1, 2, \dots, n$) равномерно стремятся к единице, так что для любого $\varepsilon > 0$ мы имеем $1 - \varphi_r(t) < \varepsilon$ ($1 \leq r \leq n$), если n достаточно велико.

Предпосылка 2° требует некоторого пояснения. Ближайший анализ показывает, что одного равномерного уменьшения интенсивности складываемых потоков, выражаемого предпосылкой 1°, еще недостаточно для того, чтобы суммарный поток приближался к простейшему; этому могут помешать скопления большого числа вызовов одного и того же потока на небольших участках — скопления, возможность которых создается тем, что последствие в каждом из складываемых потоков мы не подвергали до сих пор никаким ограничениям. Предпосылка 2° имеет целью как раз уменьшить шансы такого рода скоплений. Она говорит, что при сколь угодно большом t вероятность не получить после некоторого вызова в время t ни одного нового вызова *того же потока* должна

при $n \rightarrow \infty$ стремиться к единице равномерно по всем слагаемым потокам.

Прежде всего мы покажем, что при сделанных предположениях вероятность $V_0(t)$ отсутствия вызовов суммарного потока в промежутке $(0, t)$ приближается при $n \rightarrow \infty$ к соответствующей вероятности для простейшего потока с параметром Λ *).

Теорема Пальма. При постоянном $t > 0$ и при $n \rightarrow \infty$

$$V_0(t) \rightarrow e^{-\Lambda t}.$$

Доказательство. Формула (10.5) дает для r -го потока

$$1 - v_{0,r}(t) = \lambda_r \int_0^t \varphi_r(u) du = \lambda_r t - \lambda_r \int_0^t [1 - \varphi_r(u)] du,$$

или

$$v_{0,r}(t) = 1 - \lambda_r t + \lambda_r \int_0^t [1 - \varphi_r(u)] du;$$

поэтому в силу 2° при достаточно большом n

$$v_{0,r}(t) = 1 - \lambda_r t + \varepsilon \theta_r \lambda_r t, \quad |\theta_r| < 1, \quad 1 \leq r \leq n, \quad (14.1)$$

*) В работе Г. А. Ососкова «Одна предельная теорема для потоков однородных событий» (журнал «Теория вероятностей и ее применения», т. 1, вып. 2, 1956, 274—282) рассуждениями, в значительной степени повторяющими рассуждения А. Я. Хинчина, доказана несколько более общая теорема:

Для того чтобы суммарный поток, о котором шла речь, при $n \rightarrow \infty$ сходил к простейшему потоку с параметром Λ , достаточно выполнения условия: при фиксированном $t > 0$ и $n \rightarrow \infty$

$$\frac{1}{\Lambda} \sum_{k=1}^n \lambda_k \varphi_k(t) \rightarrow 1. \quad (1)$$

Если же параметры слагаемых потоков бесконечно малы, т. е. при любом $\varepsilon > 0$ для достаточно больших n

$$\max_{1 \leq k \leq n} \lambda_k \leq \varepsilon,$$

то условие (1) является и необходимым.

Несколько иная формулировка теоремы Г. А. Ососкова и иное доказательство даны Б. И. Григелионисом. — Б. Г.

отсюда легко находим

$$|\ln v_{0,r}(t) + \lambda_r t| < c(t) \varepsilon \lambda_r \quad (1 \leq r \leq n),$$

где $c(t) > 0$ зависит только от t . Так как в силу взаимной независимости потоков

$$V_0(t) = \prod_{r=1}^n v_{0,r}(t),$$

то отсюда

$$\begin{aligned} |\ln V_0(t) + \Lambda t| &= \left| \sum_{r=1}^n [\ln v_{0,r}(t) + \lambda_r t] \right| \leq \\ &\leq \sum_{r=1}^n |\ln v_{0,r}(t) + \lambda_r t| < c(t) \varepsilon \Lambda. \end{aligned}$$

Так как $\varepsilon > 0$ сколь угодно мало при достаточно большом n , то при $n \rightarrow \infty$

$$\ln V_0(t) \rightarrow -\Lambda t, \quad V_0(t) \rightarrow e^{-\Lambda t},$$

что и требовалось доказать.

Без достаточных оснований Пальм полагает, что доказанная теорема уже влечет за собой приближенно простейший характер суммарного потока *). Разумеется, этой теоремы еще далеко не достаточно, и мы должны теперь перейти к дальнейшему исследованию вопроса.

§ 15. Предельное поведение функций $V_k(t)$

Мы должны теперь в первую очередь убедиться, что и при любом $k > 0$ функция $V_k(t)$ нашего суммарного потока при $n \rightarrow \infty$ стремится к соответствующей функции простейшего потока с параметром Λ , т. е. к $e^{-\Lambda t} (\Lambda t)^k / k!$. С этой целью нам понадобится следующая общая

Лемма. Пусть мы имеем стационарный и ординарный поток с интенсивностью λ и функцией Пальма $\phi(t)$ и пусть $\psi(t)$, как прежде, означает вероятность поступления не менее двух вызовов за время t . Тогда при любом

*) Мы должны при этом отметить, что доказательство Пальма, с одной стороны, налагает на функции $\phi_r(t)$ излишние требования, а с другой — содержит пробелы.

$$t > 0$$

$$\psi(t) \leq \lambda t [1 - \varphi(t)].$$

Доказательство. Разобьем отрезок $(0, t)$ на m равных между собой частей (ячеек)

$$\Delta_k = \left(\frac{k-1}{m} t, \frac{k}{m} t \right) \quad (1 \leq k \leq m).$$

Поступление в отрезке $(0, t)$ по меньшей мере двух вызовов, очевидно, влечет за собой наступление по меньшей мере одного из следующих двух событий:

(A) Существует по меньшей мере одна ячейка Δ_k , содержащая не менее двух вызовов.

(B) Существует такая ячейка Δ_k ($k < m$), что как в Δ_k , так и в отрезке $\left(\frac{k}{m} t, t \right)$ содержатся вызовы.

Поэтому мы имеем

$$\psi(t) \leq P(A) + P(B). \quad (15.1)$$

Прежде всего мы имеем в силу ординарности данного потока

$$P(A) \leq m\psi\left(\frac{t}{m}\right) = t \frac{\psi\left(\frac{t}{m}\right)}{\frac{t}{m}} \rightarrow 0 \quad (m \rightarrow \infty). \quad (15.2)$$

Далее в § 9 мы обозначали через $h_0(\tau, t)$ вероятность того, что в промежутке длины τ имеются вызовы, а в следующем за ним промежутке длины t вызовов нет. $w(\tau) = h_0(\tau, t)$ есть поэтому вероятность того, что вызовы имеются как в τ , так и в t , следовательно,

$$\begin{aligned} P(B) &\leq \sum_{k=1}^{m-1} \left[w\left(\frac{t}{m}\right) - h_0\left(\frac{t}{m}, \frac{m-k}{m}t\right) \right] \leq \\ &\leq m \left[w\left(\frac{t}{m}\right) - h_0\left(\frac{t}{m}, t\right) \right] = \\ &= mw\left(\frac{t}{m}\right) \left[1 - \frac{h_0\left(\frac{t}{m}; t\right)}{w\left(\frac{t}{m}\right)} \right]. \end{aligned} \quad (15.3)$$

Так как при $m \rightarrow \infty$

$$m w\left(\frac{t}{m}\right) = t \frac{w\left(\frac{t}{m}\right)}{\frac{t}{m}} \rightarrow \lambda t,$$

то правая часть последних неравенств при $m \rightarrow \infty$ стремится к $\lambda t [1 - \varphi(t)]$. А так как $\psi(t)$ от m не зависит, то из (15.1), (15.2) и (15.3) вытекает в пределе при $m \rightarrow \infty$

$$\psi(t) \leq \lambda t [1 - \varphi(t)],$$

что и требовалось доказать.

Введем теперь следующие обозначения для событий:
 A_k — в промежутке $(0, t)$ поступает k вызовов суммарного потока;

H_1 — ни один из слагаемых потоков не дает в $(0, t)$ более одного вызова;

H_2 — по меньшей мере один из слагаемых потоков дает в $(0, t)$ более одного вызова.

Нашей целью является исследование асимптотического поведения величины $V_k(t) = P(A_k)$. Но

$$P(A_k) = P(H_1 A_k) + P(H_2 A_k),$$

и, обозначая функцию $\psi(t)$ для r -го слагаемого потока через $\psi_r(t)$, в силу доказанной леммы

$$P(H_2 A_k) \leq P(H_2) \leq \sum_{r=1}^n \psi_r(t) \leq \sum_{r=1}^n \lambda_r t [1 - \varphi_r(t)];$$

а так как в силу предпосылки 2° мы имеем при достаточно большом n

$$1 - \varphi_r(t) < \varepsilon \quad (r = 1, 2, \dots, n),$$

то при достаточно большом n

$$P(H_2 A_k) \leq \varepsilon t \Lambda,$$

и, следовательно, при $n \rightarrow \infty$

$$P(H_2 A_k) \xrightarrow{0}, \quad V_k(t) = P(H_1 A_k) + o(1). \quad (15.4)$$

Но событие $H_1 A_k$ состоит, очевидно, в том, что из n слагаемых потоков какие-то k дают в промежутке $(0, t)$ по одному вызову, тогда как остальные $n - k$ в этом проме-

жутке вызовов не дают. Поэтому, если $C(r_1, r_2, \dots, r_k)$ означает произвольное сочетание из k различных между собой чисел ряда $1, 2, \dots, n$, то

$$\begin{aligned} P(H_1 A_k) &= \sum_C \frac{v_{1r_1}(t) v_{1r_2}(t) \dots v_{1r_k}(t)}{v_{or_1}(t) v_{or_2}(t) \dots v_{or_k}(t)} \prod_{i=1}^n v_{oi}(t) = \\ &= V_0(t) \sum_C \prod_{p=1}^k \frac{v_{1r_p}(t)}{v_{or_p}(t)}, \end{aligned} \quad (15.5)$$

где суммирование производится по всем сочетаниям описанного типа.

Теперь мы можем приступить к доказательству нашего основного утверждения.

Теорема. При $n \rightarrow \infty$

$$V_k(t) \rightarrow e^{-\Delta t} \frac{(\Delta t)^k}{k!} \quad (k=0, 1, 2, \dots).$$

Доказательство. Из (14.1) следует, что при достаточно большом n

$$|w_r(t) - \lambda_r t| = |1 - v_{or}(t) - \lambda_r t| < \varepsilon \lambda_r t \quad (1 \leq r \leq n);$$

а так как $v_{1r}(t) = w_r(t) - \psi_r(t)$ и в силу доказанной леммы, при достаточно большом n , $\psi_r(t) < \varepsilon \lambda_r t$, то мы можем при постоянном t писать

$$\left. \begin{aligned} v_{1rp}(t) &= \lambda_{rp} t + q_1 \varepsilon \lambda_{rp} t, \\ v_{or_p}(t) &= 1 - \lambda_{rp} t + q_2 \varepsilon \lambda_{rp} t = 1 + q_3 \varepsilon \end{aligned} \right\} \quad (p=1, 2, \dots, k),$$

где q_1, q_2, q_3 (как и $q_4, q_5, \dots$ в дальнейшем) ограничены при $n \rightarrow \infty$. Отсюда

$$\frac{v_{1rp}(t)}{v_{or_p}(t)} = \frac{\lambda_{rp} t (1 + q_1 \varepsilon)}{1 + q_3 \varepsilon} = \lambda_{rp} t (1 + q_4 \varepsilon) \quad (p=1, 2, \dots, k),$$

и, следовательно,

$$\prod_{p=1}^k \frac{v_{1rp}(t)}{v_{or_p}(t)} = \lambda_{r_1} \lambda_{r_2} \dots \lambda_{r_k} t^k (1 + q_5 \varepsilon).$$

В силу (15.5) и теоремы Пальма (§ 14) поэтому

$$P(H_1 A_k) = e^{-\Lambda t} t^k (1 + q_e e) \sum_C \lambda_{r_1} \lambda_{r_2} \dots \lambda_{r_k}.$$

Так как в произвольно мало при достаточно большом n , то для доказательства теоремы нам в силу (15.4) достаточно убедиться, что при $n \rightarrow \infty^*$)

$$S_k = \sum_{C_k} \lambda_{r_1} \lambda_{r_2} \dots \lambda_{r_k} \rightarrow \frac{\Lambda^k}{k!}. \quad (15.6)$$

Это мы теперь и сделаем. При $k=1$ соотношение (15.6) тривиально. Пусть поэтому для некоторого $k > 1$ при $n \rightarrow \infty$

$$S_{k-1} = \sum_{C_{k-1}} \lambda_{r_1} \lambda_{r_2} \dots \lambda_{r_{k-1}} = \frac{\Lambda^{k-1}}{(k-1)!} + o(1). \quad (15.7)$$

Умножим каждый член суммы S_{k-1} на сумму всех λ_i , не входящих в него, т. е. на величину

$$\Lambda - \lambda_{r_1} - \lambda_{r_2} - \dots - \lambda_{r_{k-1}}.$$

Тогда после раскрытия всех скобок мы получим сумму произведений вида

$$\lambda_{r_1} \lambda_{r_2} \dots \lambda_{r_k},$$

где индексы $r_1, r_2, \dots, r_k$ попарно различны между собой.

Каждое такое произведение есть один из членов суммы S_k ; обратно, любой член суммы S_k будет, очевидно, получен при этой операции и притом в точности k раз [член $\lambda_{r_1} \lambda_{r_2} \dots \lambda_{r_{k-1}} \lambda_{r_k}$ получается как $(\lambda_{r_1} \lambda_{r_2} \dots \lambda_{r_{k-1}}) \lambda_{r_k}$, как $(\lambda_{r_1} \lambda_{r_2} \dots \lambda_{r_{k-2}} \lambda_{r_k}) \lambda_{r_{k-1}}$ и т. д., наконец, как $(\lambda_{r_2}, \dots, \lambda_{r_k}) \lambda_{r_1}$]. Так как, очевидно, при достаточно большом n и при любой комбинации C_{k-1}

$$\Lambda - (k-1)\varepsilon \leq \Lambda - \lambda_{r_1} - \lambda_{r_2} - \dots - \lambda_{r_{k-1}} \leq \Lambda,$$

то из нашего подсчета следует

$$(\Lambda - k\varepsilon) S_{k-1} \leq k S_k \leq \Lambda S_{k-1},$$

*) Мы здесь для большей отчетливости обозначаем различные сочетания по k из чисел $1, 2, \dots, n$ через C_k (вместо прежнего обозначения C).

и следовательно, из (15.7) при $n \rightarrow \infty$

$$(\Lambda - k\varepsilon) \left[\frac{\Lambda^{k-1}}{(k-1)!} + o(1) \right] \leq kS_k \leq \Lambda \left[\frac{\Lambda^{k-1}}{(k-1)!} + o(1) \right];$$

а так как ε при достаточно большом n как угодно мало, то $kS_k \rightarrow \Lambda^k / (k-1)!$ при $n \rightarrow \infty$, т. е.

$$\lim_{n \rightarrow \infty} S_k = \frac{\Lambda^k}{k!},$$

и наша теорема доказана индукцией.

§ 16. Предельная теорема

Только что доказанная теорема устанавливает, что в рассматриваемых нами условиях функции $V_k(t)$ для суммарного потока стремятся при $n \rightarrow \infty$ к соответствующим функциям простейшего потока с параметром Λ . Это, однако, еще не означает, что наш суммарный поток сам приближается к этому простейшему потоку. Дело в том, что, как мы видели в § 6, совокупность функций $V_k(t)$ ($k=0, 1, 2, \dots$) однозначно определяет собой данный поток лишь при условии, что это есть поток без последействия; мы же пока не рассматривали вопроса о последействии в нашем суммарном потоке. Поэтому вопрос о приближении этого суммарного потока к простейшему потоку с параметром Λ требует дальнейшего исследования.

Как показывает заключительная формула § 6, для простейшего потока с параметром Λ определяющей является формула

$$P\{x(t_i) = k_i, 1 \leq i \leq m\} = e^{-\Lambda t_m} \Lambda^{k_m} \prod_{i=1}^m \frac{(t_i - t_{i-1})^{k_i - k_{i-1} - 1}}{(k_i - k_{i-1})!}, \quad (16.1)$$

где $t_0 = k_0 = 0$, m — любое натуральное число, $0 < t_1 < t_2 < \dots < t_m$, $0 \leq k_1 \leq k_2 \leq \dots \leq k_m$ и все k_i — неотрицательные целые числа. Мы можем поэтому считать наш суммарный поток стремящимся к простейшему потоку с параметром Λ , если для этого суммарного потока вероятность, стоящая в левой части равенства (16.1), при любых m, t_i, k_i ($1 \leq i \leq m$)

и при безгранично возрастающем n имеет своим пределом правую часть этого равенства. Это мы теперь и установим.

Введем сначала более удобные для данной цели обозначения. Положим $t_i - t_{i-1} = u_i$, $k_i - k_{i-1} = l_i$, $t_m = \sum_{i=1}^m u_i = u$, $k_m = \sum_{i=1}^m l_i = k$ и обозначим через $n(u_i)$ число вызовов, поступающих в промежутке $u_i = (t_{i-1}, t_i)$. Тогда, очевидно, равенство (16.1) равносильно равенству

$$P\{n(u_i) = l_i, 1 \leq i \leq m\} = e^{-\Lambda u} \Lambda^k \prod_{i=1}^m \frac{u_i^{l_i}}{l_i!}. \quad (16.2)$$

В § 15 мы рассматривали событие $H_1 A_k$, состоящее в том, что за некоторый промежуток времени $(0, u)$ происходит k вызовов и что все эти вызовы принадлежат различным слагающим потокам. Пусть отрезок $(0, u)$ разбит на m частей $u_1, u_2, \dots, u_m$, длины которых мы будем обозначать теми же буквами, и пусть событие $H_1 A_k$ совершилось. В силу стационарности и взаимной независимости слагающих потоков для любого из поступивших в промежутке $(0, u)$ k вызовов вероятность попасть в отрезок u_i ($1 \leq i \leq m$) тогда равна u_i/u , каково бы ни было положение этого отрезка и каковы бы ни были моменты остальных поступивших вызовов. Обозначим через B событие

$$n(u_i) = l_i \quad (1 \leq i \leq m);$$

тогда из только что сказанного следует, что

$$P_{H_1 A_k}(B) = \frac{k!}{l_1! l_2! \dots l_m!} \left(\frac{u_1}{u}\right)^{l_1} \left(\frac{u_2}{u}\right)^{l_2} \dots \left(\frac{u_m}{u}\right)^{l_m}.$$

Но в § 15 мы доказали, что при $n \rightarrow \infty$

$$V_k(u) = P(A_k) \rightarrow e^{-\Lambda u} \frac{(\Lambda u)^k}{k!},$$

$$P(H_1 A_k) \rightarrow e^{-\Lambda u} \frac{(\Lambda u)^k}{k!}, \quad P(H_2 A_k) \rightarrow 0.$$

Поэтому, учитывая, что в силу $\sum_{i=1}^m l_i = k$ событие A_k есть следствие события B , мы находим при $n \rightarrow \infty$

$$\begin{aligned} P(B) &= P(A_k B) = P(H_1 A_k B) + P(H_2 A_k B) = \\ &= P(H_1 A_k) P_{H_1 A_k}(B) + P(H_2 A_k) P_{H_2 A_k}(B) \rightarrow \\ &\rightarrow e^{-\Delta u} \frac{(\Delta u)^k}{k!} \cdot \frac{k!}{l_1! l_2! \dots l_m!} \left(\frac{u_1}{u}\right)^{l_1} \left(\frac{u_2}{u}\right)^{l_2} \dots \left(\frac{u_m}{u}\right)^{l_m} = \\ &= e^{-\Delta u} \Delta^k \prod_{i=1}^m \frac{u_i^{l_i}}{l_i!}. \end{aligned}$$

Так как правая часть этого соотношения совпадает с правой частью равенства (16.2), то наша предельная теорема доказана. Мы можем, таким образом, утверждать, что при рассматриваемых нами условиях суммарный поток действительно стремится к простейшему потоку с параметром Δ .

Примечание к лемме на стр. 55. Формулировка леммы неявно предполагает, что при любом r для каждого $u > 0$ функция $\psi_r(u)$ положительна. В действительности же существуют потоки, у которых $\psi_r(u) = 0$ для всех $u \leq a_r$. Это обстоятельство было отмечено в заметке Ю. Лукашевич «Замечание об одной теореме Хинчина из теории случайных потоков», Colloquium mathematicum 7:2, 285—287, 1960. Для последующего изложения лемма Хинчина должна быть заменена на следующую: Для любого потока P и любого $r = 1, 2, \dots$

$$\lim_{u \rightarrow a_r + 0} \frac{\psi_{r+1}(u)}{\psi_r(u)} = 0,$$

где $a_r = \sup [u, \psi_r(u) = 0]$.

Доказательство этой измененной леммы является буквальным повторением рассуждений А. Я. Хинчина. — Б. Г.

ЧАСТЬ II

СИСТЕМЫ С ПОТЕРЯМИ

§ 17. Вводные замечания

После того как в первой части книги подробно изучены свойства потока поступающих вызовов, мы в последующих разделах должны рассмотреть основные вопросы, связанные с обслуживанием этого потока. Каждая станция (пункт, в который поступают вызовы) снабжена несколькими приборами, предназначенными для обслуживания вызовов; эти приборы могут быть весьма различного типа соответственно возлагаемым на них функциям; в частности, таким «прибором» может быть и человек (телефонистка, продавец в магазине, врач в амбулатории и т. п.); мы условимся ради единства терминологии во всех случаях называть обслуживающие приборы линиями. Процесс обслуживания протекает так, что всякий поступающий вызов занимает на некоторое время одну из имеющихся в момент его поступления свободных (незанятых) линий; пока линия занята каким-либо вызовом, она недоступна для вновь поступающих вызовов; период занятия какой-либо линии одним вызовом мы будем (опять-таки только ради единства терминологии) называть разговором.

Для задач обслуживания основное значение имеет различие двух типов устройства станций. Если в момент поступления какого-либо вызова имеются свободные линии, то вызов при любом устройстве станции занимает какую-либо одну из них и приступает к разговору. Различия возникают лишь в том случае, когда поступающий вызов застает все линии занятыми. При одном устройстве (системы с потерями) такой вызов просто получает отказ (или, как говорят,

«теряется»), и все дальнейшее течение процесса обслуживания идет так, как если бы этот вызов вообще не поступал; при другом устройстве (системы с ожиданием) вызов, заставший все линии занятыми, сохраняется как претендент на будущий разговор и в дальнейшем занимает одну из освободившихся линий. Эти два типа устройств отличаются друг от друга не только способом решения основных задач, но и самой их постановкой: дело в том, что уже показатели качества обслуживания в этих двух случаях совершенно различны. Для систем с потерями основным показателем является, очевидно, вероятность отказа (потери) — понятие, лишённое смысла для систем с ожиданием. Напротив, для систем с ожиданием центральной задачей служит изучение времени ожидания как случайной величины; эта задача, очевидно, не имеет никакого смысла для систем с потерями.

Ввиду столь существенных различий мы, естественно, должны рассматривать эти два типа устройств обособленно друг от друга. Системам с потерями мы посвятим вторую, а системам с ожиданием — третью часть книги. Здесь же заметим еще только, что мыслимы и системы промежуточного типа; так, например, можно заставлять неудачные вызовы ожидать, покуда число ожидающих не превзойдет определенной границы, после чего вновь поступающие вызовы уже получают отказ. Теория таких имеющих несомненное практическое значение «смешанных» систем в настоящее время еще почти не разработана.

Кроме указанного основного различия, устройства станций отличаются друг от друга еще многими другими чертами, вследствие чего число различных между собой употребительных устройств становится очень большим. Само собой понятно, что в нашей краткой монографии мы имеем возможность остановиться лишь на небольшом числе типов таких устройств и что поэтому при решении каждой задачи мы вынуждены исходить из некоторых определенных предпосылок, которые в действительности отнюдь не являются единственно возможными. Мы считаем полезным здесь же рассмотреть еще некоторые важные черты, которыми могут отличаться друг от друга различные системы обслуживания вызовов.

1° Будем предполагать, что все имеющиеся линии в равной степени доступны всем поступающим вызовам (такую

систему называют «полнодоступным пучком линий»). В противоположность этому в действительности нередко приходится встречать случаи, когда определенные категории вызовов «прикреплены» к определенным линиям и не могут занимать других линий.

2° Мы почти всегда будем предполагать поток поступающих вызовов простейшим. В части I мы сделали достаточно указаний относительно того, в какой мере это допущение соответствует действительности.

3° Назовем полнодоступный пучок упорядоченным, если его линии перенумерованы так, что поступающий вызов всегда занимает линию с наименьшим номером из числа тех, которые свободны в момент его поступления (т. е. первую, если она свободна; вторую, если первая занята, а вторая свободна, и т. д.). В неупорядоченном пучке линии занимаются в случайном порядке. На практике встречаются устройства обоих типов. Следует отметить, что для ряда основных задач вопрос об упорядоченности или неупорядоченности пучка не играет никакой роли; такова, например, вся проблема Эрланга, которой будут посвящены главы 6 и 7. Однако есть и такие задачи, которые, напротив, получают смысл только для упорядоченного пучка, такова, например, задача Пальма, которая будет рассмотрена в главе 8.

4° Для систем с ожиданием имеет во многих задачах существенное значение вопрос о порядке обслуживания ожидающих вызовов. Это обслуживание может проводиться либо в порядке очереди, либо в случайном порядке, причем в ряде задач подсчеты приводят в этих двух случаях к различным результатам.

5° Наконец, весьма важное значение для большинства задач теории обслуживания имеет вопрос о законе распределения длительности занятий (разговоров). В значительном большинстве исследований этот закон предполагается *показательным* (т. е. вероятность того, что длительность разговора будет больше t , принимается равной $e^{-\beta t}$, где $\beta > 0$ — постоянная). Этот выбор обусловлен главным образом тем, что он значительно облегчает необходимые расчеты. Можно без преувеличения сказать, что заметное большинство задач теории обслуживания решается сравнительно просто при показательном распределении длительности разговоров и, напротив, приводит к неодолимым трудностям при почти

всяком ином предположении о форме этого закона распределения.

Такое исключительное положение показательного распределения длительности разговоров обусловлено главным образом одним его важным свойством, которым мы не раз будем пользоваться в дальнейшем. Пусть $f_a(t)$ есть вероятность того, что разговор, длящийся уже a секунд, продлится еще не менее t секунд, так что при показательном распределении $f_0(t) = e^{-\beta t}$. Так как, очевидно, всегда

$$f_0(a+t) = f_0(a) f_a(t),$$

то при показательном распределении

$$e^{-\beta(a+t)} = e^{-\beta a} f_a(t),$$

откуда

$$f_a(t) = e^{-\beta t}.$$

Это означает, что *при показательном распределении длительности разговоров закон распределения оставшейся части разговора не зависит от его «возраста», т. е. от того, сколько времени он уже длится*. Именно это свойство показательного распределения в большинстве случаев и упрощает производимые расчеты. Вместе с тем оно же заставляет думать, что в практических ситуациях гипотеза показательного распределения длительности разговоров вряд ли может рассчитывать на точное осуществление и в лучшем случае способна служить лишь более или менее хорошим приближением к действительности.

Глава 6

ЗАДАЧА ЭРЛАНГА ДЛЯ КОНЕЧНОГО ПУЧКА

§ 18. Постановка задачи

В этой главе мы будем иметь дело с полнодоступным пучком (упорядоченным или нет — безразлично) из n линий, на который поступает простейший поток вызовов с параметром λ ; мы допустим, что длительность разговоров подчиняется показательному закону распределения $1 - e^{-x}$. Так как в случае общего показательного закона $1 - e^{-\beta x}$ средняя дли-

тельность разговора равна $1/\beta$, то выбор $\beta=1$ означает просто, что мы принимаем эту среднюю длительность разговора за единицу времени, что, конечно, ни в какой мере не ограничивает общности исследования.

Если известно, что в некоторый момент 0 было занято ровно k линий данного пучка ($0 \leq k \leq n$), то число $N(t)$ занятых линий в какой-либо последующий момент t есть случайная величина, значение которой определяется рядом случайных факторов: моментами окончания тех k разговоров, которые ведутся в момент 0, моментами поступления новых вызовов между 0 и t и длинами тех разговоров, которые ведутся этими вызовами. Число $N(t)$ представляет собой, таким образом, однопараметрическое семейство случайных величин, или, как говорят, *случайный процесс*. Этот процесс обладает, при сделанных нами предпосылках, одним важным свойством, позволяющим применить к его изучению хорошо разработанные методы.

Пусть $N(t_0)=i$, т. е. в момент t_0 занято i линий. Тогда последующее течение процесса в вероятностном смысле независимо от всего, что происходило до момента t_0 . В самом деле, это дальнейшее течение, как мы уже отметили, однозначно определяется следующими тремя факторами:

1. Моментами окончания тех i разговоров, которые ведутся в момент t_0 .
2. Моментами появления новых вызовов после t_0 .
3. Длительностями разговоров для вызовов, упомянутых в 2.

Но легко видеть, что ни один из этих трех случайных факторов не зависит от того, что происходило до момента t_0 . Для фактора 1 это вытекает из принятого нами показательного распределения длительности разговоров, при котором, как мы видели в § 17, длительность остающейся части разговора не зависит от его возраста. Для фактора 2 это следует из того, что поступающий поток вызовов — простейший и, следовательно, не обладает последствием. Наконец, для фактора 3 это очевидно само собой. Таким образом, действительно все три перечисленных фактора не зависят от «прошлого» нашей системы, т. е. от течения процесса до момента t_0 , а следовательно, не зависит от прошлого и течение процесса после момента t_0 , ибо оно однозначно определяется указанными тремя факторами.

Таким образом, случайный процесс $N(t)$ обладает следующим свойством: если известно $N(t_0)$, то течение процесса после момента t_0 в вероятностном смысле независимо от его течения до момента t_0 (коротко: если известно настоящее, то будущее не зависит от прошедшего). Случайные процессы, обладающие этим свойством, называют процессами Маркова.

Если в некоторый момент t занято i линий пучка [т. е. $N(t)=i$], то мы будем говорить, что в этот момент система находится в «состоянии i » ($i=0, 1, \dots, n$); всего, таким образом, для системы возможно $n+1$ различных состояний. Обозначим через $P_{ik}(t)$ ($t>0, 0 \leq i \leq n, 0 \leq k \leq n$) условную вероятность того, что система, находившаяся в некоторый момент в состоянии i , по истечении t единиц времени перейдет в состояние k [вероятность $N(a+t)=k$ при условии $N(a)=i$]. Эти «переходные» вероятности играют основную роль в исследовании процессов Маркова. Очевидно, всегда

$$P_{ik}(t) \geq 0, \quad \sum_{k=0}^n P_{ik}(t) = 1 \quad (0 \leq i \leq n, 0 \leq k \leq n).$$

Если $t_1 > 0, t_2 > 0, 0 \leq i \leq n, 0 \leq k \leq n$, то имеет место соотношение

$$P_{ik}(t_1 + t_2) = \sum_{r=0}^n P_{ir}(t_1) P_{rk}(t_2). \quad (18.1)$$

В самом деле, для того чтобы за время $t_1 + t_2$ перейти из состояния i в состояние k , система должна сначала перейти за время t_1 из состояния i в некоторое состояние r , а потом за время t_2 перейти из состояния r в состояние k , так что соотношение (18. 1) есть результат простого применения формулы полной вероятности. Очень важно отметить, что эта формула имеет место только для процессов Маркова; в самом деле, только для процессов Маркова переходную вероятность $P_{rk}(t_2)$ можно считать независимой от i ; если бы наш процесс не был процессом Маркова, то на месте $P_{rk}(t_2)$ должна была бы стоять вероятность перехода за время t_2 из состояния r в состояние k при *дополнительном условии, что система предварительно за время t_1 перешла в состояние r из состояния i* . Для процессов же Маркова вероятность $P_{rk}(t_2)$ не зависит от того, что происходило до

этого перехода, благодаря чему и имеет место формула (18.1).

Формула (18.1), иногда называемая уравнением Чэпмана — Колмогорова, лежит в основании всех исследований о процессах Маркова; в нашем изложении она также будет играть значительную роль.

Если в начальный момент 0 система находится в данном состоянии i , то вероятность заставить ее в момент $t > 0$ в состоянии k равна $P_{ik}(t)$. Мы можем, однако, сделать относительно начального момента допущение более общего характера: в момент 0 мы можем считать известным не состояние системы, а лишь начальные вероятности $P_i(0)$ ($0 \leq i \leq n$) различных состояний; этот общий случай, конечно, сводится к упомянутому нами частному случаю, когда из чисел $P_i(0)$ какое-нибудь одно равно единице (а остальные равны нулю). Вероятность $P_k(t)$ заставить систему в момент t в состоянии k по формуле полной вероятности равна

$$P_k(t) = \sum_{i=0}^n P_i(0) P_{ik}(t); \quad (18.2)$$

эта вероятность зависит как от t , так и от начальных данных $P_i(0)$ ($0 \leq i \leq n$).

Если в уравнении (18.1) помножить обе части на $P_i(0)$ и просуммировать по i от 0 до n , то в силу (18.2) мы получим

$$P_k(t_1 + t_2) = \sum_{r=0}^n P_r(t_1) P_{rk}(t_2) \quad (0 \leq k \leq n). \quad (18.3)$$

Задача Эрланга, которой мы посвятим настоящую главу состоит в отыскании вероятности заставить систему в том или ином данном состоянии. В свете изложенного нами выше такая постановка вопроса требует пояснений; случайный процесс $N(t)$ нестационарен, вероятности

$$P\{N(t) = k\} = P_k(t) \quad (0 \leq k \leq n)$$

меняются с течением времени и, кроме того, зависят еще от начальных данных, т. е. от чисел $P_i(0)$ ($0 \leq i \leq n$); представляется поэтому, что искомые в задаче Эрланга вероятности $P_k(t)$ могут быть определены лишь при данных t и $P_i(0)$ ($0 \leq i \leq n$). В приложениях, однако, обычно считают

возможным говорить о вероятности p_k заставить систему в состоянии k , независимо от выбранного момента времени и от начальных данных. Чтобы теоретически оправдать такую практику, можно попытаться установить, что процесс $N(t)$ при $t \rightarrow \infty$ безгранично приближается к некоторому стационарному процессу, не зависящему от начальных данных; говоря более конкретно, надо установить, что вероятности $P_k(t)$ при $t \rightarrow \infty$ стремятся к некоторым постоянным числам p_k ($0 \leq k \leq n$), не зависящим от начальных данных. Эти числа p_k мы тогда, естественно, и принимаем за искомые в задаче Эрланга вероятности заставить систему в том или другом определенном состоянии, ибо число p_k , с одной стороны, не зависит от начальных данных задачи, а с другой — становится сколь угодно близким к реальной вероятности $P_k(t)$, если процесс продолжается достаточно долгое время.

Итак, нашей задачей является показать, что при $t \rightarrow \infty$ функции $P_k(t)$ стремятся к числам p_k ($0 \leq k \leq n$), не зависящим от начальных данных. Разумеется, при этом числа p_k должны быть нами найдены. Для практики особо важное значение имеет число p_n — вероятность заставить все линии занятыми. Это есть вероятность потери (отказа), являющаяся для систем с потерями важнейшим показателем качества обслуживания.

Мы прежде всего редуцируем поставленную нами задачу к другой, более удобной для применения уравнения Чэпмана — Колмогорова, с помощью следующего вспомогательного предложения.

Лемма. Для того чтобы вероятности $P_k(t)$ при $t \rightarrow \infty$ стремились к не зависящим от начальных данных числам p_k ($0 \leq k \leq n$), необходимо и достаточно, чтобы к тем же пределам стремились соответственно переходные вероятности $P_{ik}(t)$ ($0 \leq k \leq n$) при любом значении i .

Доказательство. Оба утверждения леммы почти очевидны в силу (18.2).

1. Пусть $P_k(t) \rightarrow p_k$ ($t \rightarrow \infty$, $0 \leq k \leq n$), где p_k не зависит от начальных данных; выбирая тогда $P_i(0) = 1$, $P_l(0) = 0$ ($l \neq i$), мы в силу (18.2) имеем $P_k(t) = P_{ik}(t)$, и, следовательно, $P_{ik}(t) \rightarrow p_k$ ($t \rightarrow \infty$, $0 \leq i \leq n$, $0 \leq k \leq n$).

2. Пусть, обратно, $P_{ik}(t) \rightarrow p_k$ ($t \rightarrow \infty$, $0 \leq i \leq n$, $0 \leq k \leq n$); тогда в силу (18.2) при любом выборе вероят-

ностей $P_i(0)$ мы имеем

$$P_k(t) \rightarrow \sum_{i=0}^n P_i(0) p_{ik} = p_k \quad (t \rightarrow \infty, 0 \leq k \leq n),$$

так как

$$\sum_{i=0}^n P_i(0) = 1.$$

В силу этой леммы наша ближайшая задача сводится к доказательству того, что при $t \rightarrow \infty$ переходные вероятности $P_{ik}(t)$ ($0 \leq i \leq n$, $0 \leq k \leq n$) стремятся к пределам p_k , не зависящим от i .

§ 19. Теорема Маркова

Задачу, упомянутую в конце предыдущего параграфа, можно было бы пытаться решить, найдя выражения переходных вероятностей $P_{ik}(t)$ ($0 \leq i \leq n$, $0 \leq k \leq n$) для специального интересующего нас процесса $N(t)$ и стремясь затем анализом этих выражений установить существование нужных нам пределов и одновременно найти эти пределы. Однако мы предпочтем другой путь, при котором трудная задача отыскания функций $P_{ik}(t)$ может быть обойдена. В настоящем параграфе мы, не пытаясь найти переходные вероятности $P_{ik}(t)$ и их пределы, установим только самый факт существования этих пределов; такой путь становится возможным потому, что эта теорема существования представляет собой свойство очень широкого класса процессов Маркова, отнюдь не являясь характеристикой нашего специального процесса $N(t)$. После того как это будет сделано, мы в следующем параграфе, опираясь на доказанное уже существование пределов, сможем найти эти пределы для специально интересующего нас процесса, снова минуя явные выражения функций $P_{ik}(t)$.

Условимся называть процесс Маркова, характеризуемый переходными вероятностями $P_{ik}(t)$ ($0 \leq i \leq n$, $0 \leq k \leq n$), транзитивным, если существует такое $t > 0$, что $P_{ik}(t) > 0$ ($0 \leq i \leq n$, $0 \leq k \leq n$). Таким образом, для транзитивного процесса существует такой промежуток времени, в течение которого возможен переход системы из любого состояния в любое другое. Непосредственно ясно, что

интересующий нас процесс $N(t)$ транзитивен, причем в качестве t может быть выбрано любое положительное число.

Теорема Маркова*). Для любого транзитивного процесса Маркова $P_{ik}(t)$ ($0 \leq i \leq n$, $0 \leq k \leq n$) предел

$$\lim_{t \rightarrow \infty} P_{ik}(t) = p_k \quad (0 \leq i \leq n, 0 \leq k \leq n)$$

существует и не зависит от i .

Доказательство. Во всем дальнейшем k будет означать произвольное закреплённое число ряда $0, 1, \dots, n$. Положим

$$\max_{0 \leq i \leq n} P_{ik}(t) = M_k(t), \quad \min_{0 \leq i \leq n} P_{ik}(t) = m_k(t).$$

В силу (18.1) для любого i ($0 \leq i \leq n$) и любых $t > 0$, $\tau > 0$

$$P_{ik}(t + \tau) = \sum_{r=0}^n P_{ir}(\tau) P_{rk}(t) \leq M_k(t) \sum_{r=0}^n P_{ir}(\tau) = M_k(t),$$

а следовательно,

$$M_k(t + \tau) \leq M_k(t),$$

т. е. $M_k(t)$ есть невозрастающая функция от t . Подобным же образом легко убедиться, что $m_k(t)$ есть неубывающая функция от t . Отсюда следует, что $M_k(t)$ и $m_k(t)$ при $t \rightarrow \infty$ стремятся к определенным пределам. Теорема, очевидно, будет доказана, если мы покажем, что эти пределы совпадают между собой, а для этого необходимо и достаточно иметь

$$\Delta_k(t) = M_k(t) - m_k(t) \rightarrow 0 \quad (t \rightarrow \infty).$$

В дальнейшем все суммы по всем индексам будут распространяться на значения $0, 1, \dots, n$ этих индексов, вследствие чего мы можем не указывать области суммирования. Пусть $P_{ir}(t_0) > 0$ ($0 \leq i \leq n$, $0 \leq k \leq n$) (такое t_0 найдется в силу транзитивности процесса). Положим,

$$d_{il}^{(r)} = P_{ir}(t_0) - P_{il}(t_0) \quad (0 \leq i, l, r \leq n),$$

*) У Маркова эта теорема доказана для «цепей», т. е. процессов с дискретным временем; однако доказательство переносится без всяких изменений на интересующий нас случай непрерывного времени.

и будем в дальнейшем обозначать через $\sum'$ (соответственно $\sum''$) суммы, распространенные лишь на область положительных (соответственно неположительных) $d_{il}^{(r)}$. Тогда в силу

$$\sum_r P_{ir}(t_0) = \sum_r P_{ir}(t_0) = 1 \quad (0 \leq i, l \leq n)$$

мы имеем

$$0 = \sum_r d_{il}^{(r)} = \sum_r' |d_{il}^{(r)}| - \sum_r'' |d_{il}^{(r)}| \quad (0 \leq i, l \leq n),$$

или

$$\sum_r' |d_{il}^{(r)}| = \sum_r'' |d_{il}^{(r)}| = h_{il} \quad (0 \leq i, l \leq n);$$

при этом в силу $P_{lr}(t_0) > 0$ ($0 \leq l, r \leq n$)

$$h_{il} = \sum_r' d_{il}^{(r)} = \sum_r' [P_{lr}(t_0) - P_{lr}(t_0)] < \sum_r' P_{lr}(t_0) \leq \sum_r P_{lr}(t_0) = 1.$$

Это неравенство имеет место для любых i и l , вследствие чего и

$$h = \max_{0 \leq i, l \leq n} h_{il} < 1.$$

Пусть теперь q — любое натуральное число. Тогда при $0 \leq i \leq n$ $0 \leq l \leq n$ в силу (18.1)

$$\begin{aligned} P_{ik}(qt_0 + t_0) - P_{ik}(qt_0 + t_0) &= \\ &= \sum_r P_{ir}(t_0) P_{rk}(qt_0) - \sum_r P_{ir}(t_0) P_{rk}(qt_0) = \\ &= \sum_r [P_{ir}(t_0) - P_{ir}(t_0)] P_{rk}(qt_0) = \\ &= \sum_r d_{il}^{(r)} P_{rk}(qt_0) = \sum_r' d_{il}^{(r)} P_{rk}(qt_0) - \sum_r'' |d_{il}^{(r)}| P_{rk}(qt_0) \leq \\ &\leq M_k(qt_0) h_{il} - m_k(qt_0) h_{il} = h_{il} \Delta_k(qt_0) \leq h \Delta_k(qt_0). \end{aligned}$$

Так как это неравенство имеет место для любых i, l , то можно принять

$$P_{ik}(qt_0 + t_0) = M_k(qt_0 + t_0), \quad P_{ik}(qt_0 + t_0) = m_k(qt_0 + t_0);$$

тогда мы получаем

$$\Delta_k(qt_0 + t_0) \leq h \Delta_k(qt_0);$$

рекуррентное же применение этого неравенства дает

$$\Delta_k(qt_0) \leq h^{q-1} \Delta_k(t_0) \leq h^{q-1} \rightarrow 0 \quad (q \rightarrow \infty).$$

В силу монотонности функции $\Delta_k(t)$ отсюда очевидно следует, что

$$\Delta_k(t) \rightarrow 0 \quad (t \rightarrow \infty).$$

Этим теорема Маркова доказана.

§ 20. Уравнения и формулы Эрланга

Теперь мы переходим к ставшему классическим методу Эрланга определения величин p_k , существование которых нами только что доказано. В отличие от предыдущего параграфа, мы будем при этом иметь в виду исключительно наш конкретный процесс $N(t)$.

Во всем дальнейшем нам придется иметь дело с промежутком времени бесконечно малой длины τ . Условимся для краткости обозначать через $o(\tau)$ всякую бесконечно малую порядка выше τ и соединять знаком $\approx$ всякие две величины, разность которых есть величина вида $o(\tau)$.

Согласно принятым нами в § 18 предпосылкам вероятность $w(\tau)$ поступления по меньшей мере одного вызова за промежуток времени τ есть величина $\approx \lambda\tau$, а вероятность $\psi(\tau)$ поступления более одного вызова — величина вида $o(\tau)$. С другой стороны, если какая-либо линия в данный момент занята, то вероятность оставаться занятой еще в течение τ секунд (или более) для нее равна $e^{-\tau}$: если занято k линий, то вероятность того, что все они останутся занятыми в течение промежутка времени τ , равна поэтому $e^{-k\tau}$; вероятность же того, что в течение промежутка времени τ по меньшей мере одна из этих линий освободится, равна

$$1 - e^{-k\tau} \approx k\tau.$$

Поступление вызовов и освобождение линий представляют собой элементарные события, в моменты которых скачкообразно меняется величина $N(t)$. Из того, что мы до сих пор установили для вероятностей таких элементарных событий, с очевидностью следует, что вероятность наступления в промежутке длины τ по меньшей мере одного элементарного события (того или другого типа) при $\tau \rightarrow 0$ асимптотически пропорциональна τ ; вероятность же наступления в промежутке длины двух или более элементарных событий (все равно, каких типов) есть величина вида $o(\tau)$ (или, что то же, ≈ 0).

Эти замечания позволяют легко найти асимптотические выражения переходных вероятностей $P_{ik}(\tau)$ при $\tau \rightarrow 0$. Прежде всего, если $|i - k| > 1$, то переход из состояния i в состояние k требует, очевидно, наступления по меньшей мере двух элементарных событий; поэтому в силу вышесказанного при $\tau \rightarrow 0$

$$P_{ik}(\tau) \approx 0 \quad (|i - k| > 1). \quad (20.1)$$

Далее, для перехода из состояния $k < n$ в состояние $k + 1$ требуется либо наступление одного вызова, либо наступление более чем одного элементарного события; поэтому в силу вышесказанного при $\tau \rightarrow 0$

$$P_{k, k+1}(\tau) \approx \lambda \tau \quad (0 \leq k < n).$$

Чтобы система перешла из состояния $k > 0$ в состояние $k - 1$, требуется либо освобождение одной из линий, либо наступление более чем одного элементарного события; так как вероятность освобождения одной из k занятых линий за время τ при $\tau \rightarrow 0$, как мы видели выше, $\approx k\tau$, то мы находим

$$P_{k, k-1}(\tau) \approx k\tau \quad (0 < k \leq n).$$

Наконец, в силу (20.1) мы имеем при $\tau \rightarrow 0$

$$P_{kk}(\tau) \approx 1 - P_{k, k+1}(\tau) - P_{k, k-1}(\tau) \quad (0 \leq k \leq n),$$

где при $k = n$ второй, а при $k = 0$ третий член правой части надо заменить нулем; это дает

$$P_{00}(\tau) \approx 1 - \lambda \tau;$$

$$P_{kk}(\tau) \approx 1 - \lambda \tau - k\tau \quad (1 \leq k \leq n - 1);$$

$$P_{nn}(\tau) \approx 1 - n\tau.$$

Таким образом, для всех вероятностей $P_{ik}(\tau)$ нами установлены очень простые асимптотические выражения с точностью до величины вида $o(\tau)$.

Теперь мы обратимся к уравнению (18.3), в силу которого при любом постоянном $t \geq 0$

$$P_k(t + \tau) = \sum_r P_r(t) P_{rk}(\tau).$$

Применяя к вероятностям $P_{r,k}(\tau)$ в правой части этого равенства найденные нами асимптотические оценки, мы находим

$$P_0(t+\tau) = P_0(t) P_{00}(\tau) + P_1(t) P_{10}(\tau) + o(\tau) = \\ = (1 - \lambda\tau) P_0(t) + \tau P_1(t) + o(\tau);$$

$$P_k(t+\tau) = P_{k-1}(t) P_{k-1,k}(\tau) + \\ + P_k(t) P_{kk}(\tau) + P_{k+1}(t) P_{k+1,k}(\tau) + o(\tau) = \\ = \lambda\tau P_{k-1}(t) + (1 - \lambda\tau - k\tau) P_k(t) + (k+1)\tau P_{k+1}(t) + o(\tau) \\ (0 < k < n);$$

$$P_n(t+\tau) = P_{n-1}(t) P_{n-1,n}(\tau) + P_n(t) P_{nn}(\tau) + o(\tau) = \\ = \lambda\tau P_{n-1}(t) + (1 - n\tau) P_n(t) + o(\tau).$$

Отсюда

$$\frac{P_0(t+\tau) - P_0(t)}{\tau} = -\lambda P_0(t) + P_1(t) + o(1);$$

$$\frac{P_k(t+\tau) - P_k(t)}{\tau} = \lambda P_{k-1}(t) - (\lambda + k) P_k(t) + \\ + (k+1) P_{k+1}(t) + o(1) \quad (0 < k < n);$$

$$\frac{P_n(t+\tau) - P_n(t)}{\tau} = \lambda P_{n-1}(t) - n P_n(t) + o(1).$$

Если мы заставим теперь τ стремиться к нулю (сохраняя t постоянным), то убеждаемся в существовании производных всех функций $P_k(t)$ ($k=0, 1, \dots, n$) и находим в пределе

$$\left. \begin{aligned} P'_0(t) &= -\lambda P_0(t) + P_1(t); \\ P'_k(t) &= \lambda P_{k-1}(t) - (\lambda + k) P_k(t) + (k+1) P_{k+1}(t) \quad (0 < k < n); \\ P'_n(t) &= \lambda P_{n-1}(t) - n P_n(t). \end{aligned} \right\} (\mathcal{E})$$

Система $n+1$ уравнений $(\mathcal{E})$ с $n+1$ неизвестными функциями $P_k(t)$ ($k=0, 1, \dots, n$) называется системой Эрланга. Так как все уравнения этой системы однородны, то искомые функции содержат произвольный постоянный множитель, который может быть определен из очевидного «нормировочного» условия

$$\sum_{k=0}^n P_k(t) = 1.$$

Как мы уже говорили, нам нет надобности искать решения системы дифференциальных уравнений $(\mathcal{E})$. В § 19 мы

доказали, что для любого k ($0 \leq k \leq n$) существует предел

$$\lim_{t \rightarrow \infty} P_k(t) = p_k.$$

Отсюда следует, что правые части всех уравнений (8) при $t \rightarrow \infty$ имеют пределы. Переходя к левым частям, мы видим, что все производные $P'_k(t)$ при $t \rightarrow \infty$ стремятся к пределам; но такой предел может быть только нулем, так как, если бы какое-нибудь $P'_k(t)$ стремилось к числу, отличному от нуля, соответствующее $|P_k(t)|$ при $t \rightarrow \infty$ возрастало бы безгранично, что (независимо от реального смысла величин $P_k(t)$ как вероятностей) невозможно уже в силу теоремы Маркова. Таким образом, мы приходим к выводу, что

$$P'_k(t) \rightarrow 0 \quad (t \rightarrow \infty) \quad (0 \leq k \leq n),$$

вследствие чего система (8) в пределе при $t \rightarrow \infty$ дает

$$\left. \begin{aligned} -\lambda p_0 + p_1 &= 0; \\ \lambda p_{k-1} - (\lambda + k)p_k + (k+1)p_{k+1} &= 0 \quad (0 < k < n); \\ \lambda p_{n-1} - np_n &= 0. \end{aligned} \right\} \quad (20.2)$$

Эта простая система линейных уравнений вместе с нормировочным условием $\sum_{k=0}^n p_k = 1$ может служить для однозначного определения искомых чисел p_k .

Если положить

$$\lambda p_{k-1} - kp_k = z_k \quad (1 \leq k \leq n),$$

система (20.2) может быть записана в виде

$$z_1 = 0, \quad z_k - z_{k+1} = 0 \quad (0 < k < n), \quad z_n = 0,$$

откуда $z_k = 0$ ($1 \leq k \leq n$); это же дает

$$p_k = \frac{\lambda}{k} p_{k-1} \quad (1 \leq k \leq n),$$

и, следовательно,

$$p_k = \frac{\lambda^k}{k!} p_0 \quad (1 \leq k \leq n).$$

Применяя нормировочное условие, находим

$$p_0 = \frac{1}{\sum_{l=0}^n \frac{\lambda^l}{l!}},$$

и, следовательно,

$$p_k = \frac{\frac{\lambda^k}{k!}}{\sum_{l=0}^n \frac{\lambda^l}{l!}} \quad (k=0, 1, \dots, n). \quad (20.3)$$

Формулы (20.3), называемые обычно формулами Эрланга, полностью решают поставленную нами задачу. В частности, вероятность «потери» дается формулой

$$p_n = \frac{\frac{\lambda^n}{n!}}{\sum_{l=0}^n \frac{\lambda^l}{l!}}.$$

Полезно отметить, насколько решающую роль во всем проведенном исследовании играла предпосылка о показательном распределении длительности разговоров; только при этом допущении процесс $N(t)$ становится процессом Маркова; при отказе от этого допущения все развитые нами в § 19 и 20 методы становятся принципиально неприменимыми. В специальной литературе имеется целый ряд попыток доказать, что формулы Эрланга остаются в силе и при любом другом распределении длительности разговоров. Однако, насколько мы видим, эти попытки не привели до сих пор к сколько-нибудь законченным результатам*).

*) Вскоре после опубликования этой работы сам А. Я. Хинчин (см. «О формулах Эрланга в теории массового обслуживания», стр. 199) и Б. А. Севастьянов («Эргодическая теорема для марковских процессов и ее приложение к телефонным задачам с отказами», Теор. вероят., т. 2, вып. 1, 1957, стр. 106 — 116) доказали это утверждение.— Б. Г.

§ 21. Эргодическая теорема

Любая вероятность в любой теории получает реальный смысл лишь в том случае, если известна реальная совокупность объектов, в которой эта вероятность интерпретирует собой долю того или другого признака. Каков же реальный смысл тех вероятностей p_k , которыми мы занимались в последних параграфах? Что означает, в частности, «вероятность потери» p_n ?

В подавляющем большинстве специальных исследований реальная интерпретация этих вероятностей носит один и тот же вполне определенный характер: их истолковывают как средние относительные времена пребывания системы в соответствующих состояниях. Это означает следующее. Обозначим через $x_k(t)$ величину, равную 1, если система в момент t находится в состоянии k , и равную 0 в противном случае. Тогда интеграл

$$\int_0^T x_k(t) dt$$

представляет собой суммарную длину тех промежутков времени (между 0 и T), в течение которых система находится в состоянии k , а отношение

$$\frac{1}{T} \int_0^T x_k(t) dt$$

дает нам среднее относительное время пребывания системы в состоянии k (за промежуток времени $(0, T)$). Под вероятностью p_k заставить систему в состоянии k понимают тогда предел

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x_k(t) dt. \quad (21.1)$$

Представляется, однако, очевидным, что определенные нами в предшествующих параграфах величины p_k непосредственно не допускают подобного истолкования. Это видно уже из того, что величина $x_k(t)$ представляет собой с принятой нами точки зрения случайную функцию (случайный

процесс), а следовательно, предел (21.1) (если он существует)—случайную величину, которая поэтому не может отождествляться с вероятностью p_k , по сущности своей, не зависящей от случая. С другой стороны, вероятности p_k нами определены как пределы при $t \rightarrow \infty$ вероятностей $P_k(t)$; если все вероятности понимать как средние времена пребывания системы в том или ином состоянии, то величины $P_k(t)$ не допускают, как легко видеть, никакой разумной интерпретации.

Таким образом, избранное нами определение вероятностей p_k , а также и развитый нами метод их вычисления непосредственно не дают никаких оснований для отождествления их с пределами вида (21.1) вопреки установившейся во всей прикладной литературе практике. Если это отождествление невозможно, то все же с практической точки зрения представляется весьма желательным найти соображения, позволяющие с известным основанием считать интегралы

$$X_k(t) = \frac{1}{T} \int_0^T x_k(t) dt$$

при больших значениях T хотя бы приближенно совпадающими с определенными нами вероятностями p_k ; если бы это удалось, то такое сближение в значительной степени оправдало бы общепринятую в специальной литературе практику понимания вероятностей p_k как средних относительных времен пребывания — практику, очень удобную в прикладных задачах.

Так как $X_k(t)$ есть с нашей точки зрения случайная величина, то близость ее к не зависящей от случая величине p_k при самых благоприятных обстоятельствах может утверждаться лишь с некоторой (достаточно большой) вероятностью. В настоящем параграфе мы докажем предложение, идущее в указанном направлении так далеко, как только можно было бы надеяться; именно, имеет место

Теорема. *Как бы мало ни было $\varepsilon > 0$,*

$$\lim_{T \rightarrow \infty} P \{ |X_k(t) - p_k| > \varepsilon \} = 0.$$

Предварительное замечание. Предложения подобного рода, в которых вероятность некоторого состояния системы, первоначально определенная как доля в большой

совокупности систем одинакового строения, сближается затем со средним временем пребывания в данном состоянии для какой-либо одной системы за большой промежуток времени, в теоретической физике называются обычно эргодическими теоремами. Доказываемое нами предложение представляет собой весьма типичный пример эргодической теоремы.

Доказательство. Во всем дальнейшем начальные данные (вероятности $P_k(0)$, $0 \leq k \leq n$) предполагаются произвольными, но твердо установленными. Математическое ожидание случайной величины ξ мы будем обозначать через $M\xi$ или $M\{\xi\}$. Так как величина $x_k(t)$ может принимать значения 1 и 0 с соответственными вероятностями $P_k(t)$ и $1 - P_k(t)$, то

$$Mx_k(t) = P_k(t). \quad (21.2)$$

Так как $P_k(t) \rightarrow p_k$ при $t \rightarrow \infty$, то

$$\int_0^T [P_k(t) - p_k] dt = o(T) \quad (T \rightarrow \infty).$$

Поэтому при достаточно большом T

$$\begin{aligned} P\{|X_k(T) - p_k| > \varepsilon\} &= P\left\{\left|\int_0^T [x_k(t) - p_k] dt\right| > \varepsilon T\right\} < \\ &< P\left\{\left|\int_0^T [x_k(t) - P_k(t)] dt\right| > \frac{\varepsilon}{2} T\right\}. \end{aligned}$$

Отсюда, применяя неравенство Чебышева, находим в силу (21.2)

$$\begin{aligned} P\{|X_k(T) - p_k| > \varepsilon\} &< \frac{4}{\varepsilon^2 T^2} M\left\{\left[\int_0^T [x_k(t) - P_k(t)] dt\right]^2\right\} = \\ &= \frac{4}{\varepsilon^2 T^2} M\left\{\int_0^T \int_0^T [x_k(u) - P_k(u)][x_k(v) - P_k(v)] du dv\right\} = \\ &= \frac{4}{\varepsilon^2 T^2} \int_0^T \int_0^T [M\{x_k(u)x_k(v)\} - P_k(u)P_k(v)] du dv = \\ &= \frac{8}{\varepsilon^2 T^2} \int_{0 \leq u \leq v \leq T} [M\{x_k(u)x_k(v)\} - P_k(u)P_k(v)] du dv. \end{aligned}$$

Так как величина $x_k(u)x_k(v)$ может принимать лишь значения 1 и 0, то $M\{x_k(u)x_k(v)\} = P\{x_k(u)x_k(v) = 1\}$ есть вероятность застать систему в состоянии k как в момент u , так и в момент v ; поэтому при $v > u$

$$M\{x_k(u)x_k(v)\} = P_k(u)P_{kk}(v-u),$$

и мы находим

$$\begin{aligned} P\{|X_k(T) - p_k| > \varepsilon\} &< \frac{8}{\varepsilon^2 T^2} \iint_{0 \leq u \leq v \leq T} P_k(u)[P_{kk}(v-u) - \\ &- P_k(v)] du dv = \\ &= \frac{8}{\varepsilon^2 T^2} \int_0^T P_k(u) du \int_u^T [P_{kk}(v-u) - P_k(v)] dv = \\ &= \frac{8}{\varepsilon^2 T^2} \int_0^T P_k(u) du \int_0^{T-u} [P_{kk}(z) - P_k(z+u)] dz. \end{aligned}$$

Пусть теперь $\delta > 0$ произвольно мало и A настолько велико, что при $z > A$

$$|P_{ik}(z) - p_k| < \frac{\delta}{2} \quad (0 \leq i \leq n).$$

Тогда при $z > A$, $u > 0$

$$\begin{aligned} |P_{kk}(z) - p_k| &< \frac{\delta}{2}, \\ |P_k(z+u) - p_k| &= \left| \sum_i P_i(0)P_{ik}(z+u) - p_k \right| = \\ &= \left| \sum_i P_i(0)[P_{ik}(z+u) - p_k] \right| \leq \frac{\delta}{2} \sum_i P_i(0) = \frac{\delta}{2}, \end{aligned}$$

и следовательно,

$$|P_{kk}(z) - P_k(z+u)| < \delta.$$

Поэтому при $T > A$

$$\begin{aligned} P\{|X_k(T) - p_k| > \varepsilon\} &\leq \\ &\leq \frac{8}{\varepsilon^2 T^2} \int_0^T du \int_0^T |P_{kk}(z) - P_k(z+u)| dz \leq \\ &\leq \frac{8}{\varepsilon^2 T^2} \int_0^T du \left\{ \int_0^A dz + \int_A^T \delta dz \right\} \leq \frac{8}{\varepsilon^2 T^2} \{AT + \delta T^2\} = \frac{8A}{\varepsilon^2 T} + \frac{8\delta}{\varepsilon^2}. \end{aligned}$$

Беря сначала δ достаточно малым, выбирая затем A описанным выше образом и беря, наконец, T достаточно большим, мы видим, что правая часть последнего неравенства сколь угодно мала при достаточно большом T . Этим наша теорема доказана.

Глава 7

ЗАДАЧА ЭРЛАНГА ДЛЯ БЕСКОНЕЧНОГО ПУЧКА

§ 22. Уравнение для производящей функции

Если число линий в пучке бесконечно, то рассматриваемую установку нельзя уже причислять к «системам с потерями», так как потери становятся невозможными. Расчет вероятностей различных состояний сохраняет, однако, практическое значение и в этом случае, так как на практике встречаются такие положения, когда потери недопустимы и число линий должно быть достаточно большим для того, чтобы вероятность потери оказалась пренебрегаемо малой; в таких случаях вероятности различных состояний дают возможность оценить степень использования системы, что в свою очередь имеет значение для расчета быстроты износа и других экономических показателей.

Если в формулах Эрланга (20.3) перейти к пределу при $n \rightarrow \infty$, то мы получаем

$$p_k = e^{-\lambda} \frac{\lambda^k}{k!} \quad (k=0, 1, 2, \dots);$$

можно поэтому предвидеть, что это пуассоновское распределение и даст нам вероятности различных состояний в случае бесконечного пучка. Однако метод, которым мы пришли к формулам (20.3), в случае бесконечного пучка оказывается неприменимым, так как теорема Маркова, на которой он основан, существенным образом предполагает число состояний конечным. Мы покажем, что случай бесконечного пучка может быть очень просто изучен методом производящих функций.

Те рассуждения, которые привели нас в § 20 к системе (8) уравнений Эрланга, сохраняются почти полностью и в случае бесконечного пучка. Различие состоит, очевидно, лишь в том, что группа уравнений системы (8), построенная

нами в § 20 для $0 < k < n$, теперь имеет место для любого $k > 0$, последнее же из уравнений системы (6) отпадает совсем. Таким образом, мы получаем для вероятностей различных состояний в случае бесконечного пучка систему уравнений

$$\left. \begin{aligned} P'_0(t) &= -\lambda P_0(t) + P_1(t); \\ P'_k(t) &= \lambda P_{k-1}(t) - (\lambda + k) P_k(t) + (k+1) P_{k+1}(t) \end{aligned} \right\} \quad (k > 0). \quad (6^*)$$

Эту систему можно считать более простой, чем система (6), так как здесь для всех $k > 0$ мы имеем уравнения одинакового типа; именно это обстоятельство и позволяет применить к решению системы (6^{*}) метод производящих функций (к системе (6) он непосредственно применен быть не может).

Положим

$$\Phi(t, x) = \sum_{k=0}^{\infty} P_k(t) x^k$$

[ряд, очевидно, сходится при $|x| \leq 1$ и любом t]; положим еще для $k > 0$

$$\lambda P_{k-1}(t) - k P_k(t) = Q_k(t);$$

тогда система уравнений (6^{*}) получает более краткий вид:

$$\left. \begin{aligned} P'_0(t) &= -Q_1(t); \\ P'_k(t) &= Q_k(t) - Q_{k+1}(t) \quad (k > 0). \end{aligned} \right\}$$

Отсюда

$$\begin{aligned} \frac{\partial \Phi}{\partial t} &= \sum_{k=0}^{\infty} P'_k(t) x^k = -Q_1(t) + \sum_{k=1}^{\infty} [Q_k(t) - Q_{k+1}(t)] x^k = \\ &= (x-1) \sum_{k=1}^{\infty} Q_k(t) x^{k-1} = \\ &= (x-1) \left\{ \lambda \sum_{k=1}^{\infty} P_{k-1}(t) x^{k-1} - \sum_{k=1}^{\infty} k P_k(t) x^{k-1} \right\} = \\ &= (x-1) \left\{ \lambda \Phi - \frac{\partial \Phi}{\partial x} \right\}, \end{aligned}$$

или

$$\frac{\partial \Phi}{\partial t} + (x-1) \frac{\partial \Phi}{\partial x} - \lambda (x-1) \Phi = 0. \quad (22.1)$$

Это простое уравнение с частными производными первого порядка и послужит нам для определения функции $\Phi(t, x)$. Прежде всего мы его еще несколько упростим преобразованием неизвестной функции. Положим

$$e^{\lambda(x-1)(1-e^{-t})} = G(t, x)$$

и

$$\Phi(t, x) = G(t, x) F(t, x),$$

где $F(t, x)$ — новая неизвестная функция. Мы будем иметь

$$\frac{\partial \Phi}{\partial t} = G \frac{\partial F}{\partial t} + FG\lambda(x-1)e^{-t};$$

$$\frac{\partial \Phi}{\partial x} = G \frac{\partial F}{\partial x} + FG\lambda(1-e^{-t}),$$

вследствие чего

$$\begin{aligned} \frac{\partial \Phi}{\partial t} + (x-1) \frac{\partial \Phi}{\partial x} - \lambda(x-1)\Phi &= G \left\{ \frac{\partial F}{\partial t} + \lambda(x-1)e^{-t}F + \right. \\ &+ (x-1) \frac{\partial F}{\partial x} + \lambda(x-1)(1-e^{-t})F - \lambda(x-1)F \Big\} = \\ &= G \left\{ \frac{\partial F}{\partial t} + (x-1) \frac{\partial F}{\partial x} \right\}, \end{aligned}$$

и уравнение (22.1) равносильно уравнению

$$\frac{\partial F}{\partial t} + (x-1) \frac{\partial F}{\partial x} = 0. \quad (22.2)$$

Положим теперь

$$(x-1)e^{-t} = L(t, x)$$

и составим функциональный определитель

$$\frac{D(F, L)}{D(t, x)} = e^{-t} \left\{ \frac{\partial F}{\partial t} + (x-1) \frac{\partial F}{\partial x} \right\};$$

уравнение (22.2) равносильно уравнению

$$\frac{D(F, L)}{D(t, x)} = 0,$$

и значит, общим решением его служит соотношение

$$F = R(L) = R[(x-1)e^{-t}],$$

где R — произвольная дифференцируемая функция своего аргумента. Для искомой функции Φ мы отсюда находим выражение

$$\Phi(t, x) = e^{\lambda(x-1)(1-e^{-t})} R[(x-1)e^{-t}]. \quad (22.3)$$

Это — общее решение уравнения (22.1). Для решения нашей задачи мы должны с помощью начальных данных определить вид функции R .

§ 23. Решение задачи

Пусть в начальный момент $t=0$ мы имеем

$$P_k(0) = a_k \quad (k=0, 1, \dots);$$

в частности, если известно, что при $t=0$ система находится в состоянии i , то $a_i=1$, $a_k=0$ ($k \neq i$); во всех случаях $\sum_{k=0}^{\infty} a_k = 1$. Полагая тогда $t=0$ в выражении (22.3) функции $\Phi(t, x)$, мы находим

$$\Phi(0, x) = \sum_{k=0}^{\infty} P_k(0) x^k = \sum_{k=0}^{\infty} a_k x^k = R(x-1),$$

откуда

$$R(z) = \sum_{k=0}^{\infty} a_k (z+1)^k.$$

В частности,

$$R[(x-1)e^{-t}] = \sum_{k=0}^{\infty} a_k [1 + (x-1)e^{-t}]^k, \quad (23.1)$$

и для производящей функции мы получаем уже окончательное выражение

$$\Phi(t, x) = e^{\lambda(x-1)(1-e^{-t})} \sum_{k=0}^{\infty} a_k [1 + (x-1)e^{-t}]^k *).$$

Наша задача состоит в том, чтобы, пользуясь этим выражением, найти пределы вероятностей $P_k(t)$ при $t \rightarrow \infty$.

Как показывает соотношение (23.1), величина $R[(x-1)e^{-t}]$ может быть представлена в виде (сходящегося при $t \geq 0$,

*) Этот ряд заведомо сходится при $t \geq 0$, $|x| \leq 1$. В самом деле, легко убедиться, что при этих условиях и $|1 + (x-1)e^{-t}| \leq 1$.

$|x| \leq 1$) ряда по степеням x :

$$R[(x-1)e^{-t}] = \sum_{k=0}^{\infty} b_k(t) x^k.$$

Убедимся теперь, что при $t \rightarrow \infty$ мы имеем

$$b_0(t) \rightarrow 1, \quad b_k(t) \rightarrow 0 \quad (k > 0).$$

В самом деле, в силу (23.1)

$$\sum_{k=0}^{\infty} b_k(t) x^k = R[(x-1)e^{-t}] = \sum_{k=0}^{\infty} a_k (xe^{-t} + 1 - e^{-t})^k; \quad (23.2)$$

так как $a_k \geq 0$, $e^{-t} > 0$ и $1 - e^{-t} \geq 0$, то сравнение левой части с правой прежде всего показывает, что все $b_k(t) \geq 0$. Далее, полагая в этом равенстве $x=0$, мы находим

$$b_0(t) = \sum_{k=0}^{\infty} a_k (1 - e^{-t})^k,$$

откуда при $t \rightarrow \infty$

$$b_0(t) \rightarrow \sum_{k=0}^{\infty} a_k = 1;$$

наконец, полагая в (23.2) $x=1$, мы находим при любом $t \geq 0$

$$\sum_{k=0}^{\infty} b_k(t) = \sum_{k=0}^{\infty} a_k = 1,$$

откуда в силу предыдущего $b_k(t) \rightarrow 0$ ($k > 0$, $t \rightarrow \infty$), и наше утверждение доказано.

В силу (22.3) мы теперь имеем

$$\begin{aligned} \Phi(t, x) &= \sum_{n=0}^{\infty} P_n(t) x^n = R[(x-1)e^{-t}] e^{\lambda(x-1)(1-e^{-t})} = \\ &= \sum_{k=0}^{\infty} b_k(t) x^k e^{-\lambda(1-e^{-t})} \sum_{r=0}^{\infty} \frac{\lambda^r (1-e^{-t})^r}{r!} x^r, \end{aligned}$$

откуда

$$P_n(t) = e^{-\lambda(1-e^{-t})} \sum_{k=0}^n b_k(t) \frac{\lambda^{n-k} (1-e^{-t})^{n-k}}{(n-k)!} \quad (n \geq 0).$$

Если теперь при фиксированном n заставить t безгранично возрастать, то, как мы видели выше, $b_0(t) \rightarrow 1$, $b_k(t) \rightarrow 0$ ($k > 0$), а потому

$$P_n(t) \rightarrow e^{-\lambda} \frac{\lambda^n}{n!} \quad (t \rightarrow \infty, n = 0, 1, 2, \dots),$$

что полностью решает поставленную задачу. Мы приходим при этом к тем самым формулам, которые мы в начале § 22 в порядке гипотезы получили предельным переходом из формул Эрланга для конечного пучка.

§ 24. Поток с переменным параметром

Методы, изложенные нами в §§ 22 и 23, позволяют легко решить задачу Эрланга для бесконечного пучка и в том случае, когда параметр поступающего потока вызовов меняется с течением времени. Мы уже имели дело с этим случаем в § 5. Будем в дальнейшем и здесь обозначать через $\lambda(t)$ «мгновенное значение» параметра в момент t , определяемое формулой (5.1).

Так как уравнения системы (6*) § 22 имеют чисто локальную природу (относятся к некоторому определенному моменту времени t), то проведенный нами вывод этих уравнений остается в полной силе и в случае переменного параметра, только на место постоянного λ становится «мгновенное значение» $\lambda(t)$ этого параметра, вообще говоря, различное в различные моменты времени t . Таким образом, в качестве исходной системы уравнений для определения функций $P_k(t)$ мы имеем

$$\left. \begin{aligned} P'_0(t) &= -\lambda(t) P_0(t) + P_1(t); \\ P'_k(t) &= \lambda(t) P_{k-1}(t) - [\lambda(t) + k] P_k(t) + \\ &\quad + (k+1) P_{k+1}(t) \quad (k > 0). \end{aligned} \right\} \quad (24.1)$$

Полагая

$$\Phi(t, x) = \sum_{k=0}^{\infty} P_k(t) x^k,$$

мы из системы (24.1) в точности как в § 22 находим для производящей функции $\Phi(t, x)$ уравнение с частными производными

$$\frac{\partial \Phi}{\partial t} + (x-1) \frac{\partial \Phi}{\partial x} - \lambda(t) (x-1) \Phi = 0, \quad (24.2)$$

отличающееся от уравнения (22.1) только тем, что на месте постоянного параметра λ стоит теперь функция $\lambda(t)$. Это отличие, несущественное для вывода уравнения (24.2), в значительной мере влияет на его решение, заставляя нас обратиться к другой замене искомой функции.

Положим во всем дальнейшем

$$e^{-t} \int_0^t e^{u\lambda(u)} du = \Lambda(t);$$

$$e^{(x-1)\Lambda(t)} = G(t, x);$$

$$\Phi(t, x) = G(t, x) F(t, x),$$

где $F(t, x)$ — новая неизвестная функция. Производя эту замену в (24.2), мы легко находим для функции $F(t, x)$ уравнение

$$\frac{\partial F}{\partial t} + (x-1) \frac{\partial F}{\partial x} = 0,$$

в точности совпадающее с уравнением (22.2). В § 22 мы видели, что общим решением этого уравнения служит

$$F(t, x) = R[(x-1)e^{-t}],$$

где R — произвольная дифференцируемая функция своего аргумента. Отсюда

$$\Phi(t, x) = e^{(x-1)\Lambda(t)} R[(x-1)e^{-t}]. \quad (24.3)$$

Теперь мы должны определить вид функции R с помощью начальных данных $P_k(0) = a_k$ ($k=0, 1, 2, \dots$). Мы находим, как в § 23,

$$R(z) = \sum_{k=0}^{\infty} a_k (z+1)^k,$$

откуда, в частности,

$$R[(x-1)e^{-t}] = \sum_{k=0}^{\infty} a_k [1 + (x-1)e^{-t}]^k, \quad (24.4)$$

где ряд заведомо сходится при $t \geq 0$, $|x| \leq 1$. Отсюда с помощью (24.3) мы получаем окончательное выражение производящей функции

$$\Phi(t, x) = e^{(x-1)\Lambda(t)} \sum_{k=0}^{\infty} a_k [1 + (x-1)e^{-t}]^k. \quad (24.5)$$

В § 23 мы полагали

$$R[(x-1)e^{-t}] = \sum_{k=0}^{\infty} b_k(t) x^k, \quad (24.6)$$

и убедились, что при $t \rightarrow \infty$

$$b_0(t) \rightarrow 1, \quad b_k(t) \rightarrow 0, \quad (k > 0).$$

Эти выводы сохраняют силу и в нашем новом случае, так как новая функция $R(z)$ ничем не отличается от прежней.

Мы имеем в силу (24.5), (24.4) и (24.6)

$$\begin{aligned} \Phi(t, x) &= \sum_{n=0}^{\infty} P_n(t) x^n = e^{(x-1)\Lambda(t)} \sum_{k=0}^{\infty} b_k(t) x^k = \\ &= e^{-\Lambda(t)} \sum_{r=0}^{\infty} \frac{[\Lambda(t)]^r}{r!} x^r \sum_{k=0}^{\infty} b_k(t) x^k = \\ &= e^{-\Lambda(t)} \sum_{n=0}^{\infty} \left\{ \sum_{r=0}^n \frac{[\Lambda(t)]^r}{r!} b_{n-r}(t) \right\} x^n, \end{aligned}$$

и следовательно,

$$P_n(t) = e^{-\Lambda(t)} \sum_{r=0}^n \frac{[\Lambda(t)]^r}{r!} b_{n-r}(t) \quad (n=0, 1, 2, \dots).$$

Допустим теперь, что при $t \rightarrow \infty$ параметр $\lambda(t)$ остается ограниченным;

$$\lambda(t) < \Lambda \quad (t \geq 0).$$

Тогда и

$$\Lambda(t) = e^{-t} \int_0^t e^u \lambda(u) du < \Lambda \quad (t \geq 0);$$

поэтому для любого постоянного n при $t \rightarrow \infty$ в сумме

$$\sum_{r=0}^n \frac{[\Lambda(t)]^r}{r!} b_{n-r}(t)$$

все члены, кроме последнего ($r=n$), стремятся к нулю, в то время как последний бесконечно мало отличается от $[\Lambda(t)]^n/n!$. Поэтому при $t \rightarrow \infty$ и постоянном n

$$P_n(t) - e^{-\Lambda(t)} \frac{[\Lambda(t)]^n}{n!} \rightarrow 0,$$

т. е. закон распределения $P_n(t)$ безгранично приближается к закону Пуассона с (переменным) параметром

$$\Lambda(t) = e^{-t} \int_0^t e^{u\lambda} (u) du.$$

Этим поставленная задача решена.

§ 25. Бесконечный пучок при произвольном законе распределения длин разговоров

В §§ 22—24 мы убедились, что для задачи Эрланга бесконечный пучок в известном смысле представляет собой более простой объект исследования, чем конечный (в частности, в § 24 мы до конца рассмотрели случай входящего потока с переменным параметром — задача, которая, насколько нам известно, еще не решена для конечного пучка). В нашем изложении сравнительная простота трактовки случая бесконечного пучка все время связывалась с возможностью применения метода производящих функций; при этом мы ради соблюдения самой тесной аналогии с теорией конечного пучка исходили всегда из системы дифференциальных уравнений Эрланга. На самом деле, задача Эрланга для бесконечного пучка представляет собой чрезвычайно простую проблему, которая может быть решена и вполне элементарными средствами; при этом оказывается, что показательный закон распределения длин разговоров, служивший важной предпосылкой в методе уравнений Эрланга, при элементарной трактовке задачи без существенных усложнений может быть заменен любым другим законом. Ради этого важного обобщения мы и остановимся в настоящем параграфе на элементарном выводе формул Эрланга для бесконечного пучка.

Мы сохраним все предпосылки § 22 с той разницей, что вероятность $F(x)$ для наудачу выбранного разговора иметь длину $> x$ мы будем предполагать произвольной невозрастающей функцией, подчиненной только требованиям

$$F(0) = 1, \quad F(+\infty) = 0, \quad - \int_0^{\infty} x dF(x) = \int_0^{\infty} F(x) dx = 1,$$

из которых последнее выражает собой соглашение принимать за единицу времени среднюю длительность разговора. Будем, как прежде, обозначать через $P_k(t)$ вероятность того, что в момент $t > 0$ ведется k разговоров (или, что то же, занято k линий). Наша задача — показать, что

$$\lim_{t \rightarrow \infty} P_k(t) = e^{-\lambda} \frac{\lambda^k}{k!} \quad (k = 0, 1, 2, \dots).$$

Пусть $X = (x_1, x_2, \dots, x_r)$ — произвольный r -мерный вектор, принадлежащий области D_r ($0 < x_1 < x_2 < \dots < x_r < t$) r -мерного пространства. Условимся называть гипотезой (r, X) событие, состоящее в том, что в промежутке $(0, t)$ происходит r вызовов и моменты t_i ($0 < t_1 < t_2 < \dots < t_r < t$) этих вызовов удовлетворяют неравенствам

$$x_i < t_i < x_i + dx_i \quad (1 \leq i \leq r).$$

Так как поток вызовов — простейший с параметром λ , то вероятность гипотезы (r, X) с точностью до бесконечно малых высших порядков выражается при малых dx_i формулой

$$P(r, X) =$$

$$= \lambda e^{-\lambda x_1} dx_1 \lambda e^{-\lambda(x_2 - x_1)} dx_2 \dots \lambda e^{-\lambda(x_r - x_{r-1})} dx_r e^{-\lambda(t - x_r)},$$

где последний множитель есть вероятность того, что между моментами x_r и t вызовов не поступает. Отсюда

$$P(r, X) = e^{-\lambda t} \lambda^r dx_1 dx_2 \dots dx_r, \quad (25.1)$$

(и, в частности, не зависит от вектора X). Пусть $P_{k, (r, X)}(t)$ — условная вероятность застать k разговоров в момент t , если имеет место гипотеза (r, X) . Очевидно, $P_{k, (r, X)}(t)$ может быть отличной от нуля лишь при $r \geq k$. Так как для разговора, начавшегося в момент x ($0 < x < t$), вероятность не закончиться к моменту t равна $F(t - x)$, то

$$P_{k, (r, X)}(t) = \sum_C \prod_{i=1}^k F(t - x_{s_i}) \prod_{\substack{s=1 \\ s \neq s_i}}^r [1 - F(t - x_s)], \quad (25.2)$$

где C — любое сочетание $(s_1, s_2, \dots, s_k)$ по k чисел ряда $(1, 2, \dots, r)$ и где суммирование производится по всем таким сочетаниям.

По формуле полной вероятности мы имеем

$$P_k(t) = \sum_{r, X} P_r(r, X) P_{k, (r, X)}(t),$$

где по r мы должны суммировать от k до ∞ , а по X интегрировать по области D_r . Поэтому мы находим в силу (25.1) и (25.2)

$$P_k(t) = e^{-\lambda t} \sum_{r=k}^{\infty} \frac{\lambda^r}{r!} \int_{D_r} \sum_C \prod_{i=1}^k F(t - x_{s_i}) \prod_{\substack{s=1 \\ s \neq s_i}}^r [1 - F(t - x_s)] dx_1 \dots dx_r.$$

Подынтегральная функция здесь, очевидно, симметрична относительно переменных $x_1, x_2, \dots, x_r$. Интеграл поэтому не изменится, если в определении области D_r мы заменим порядок $x_1 < x_2 < \dots < x_r$ переменных интегриации любым другим порядком; а так как всех таких порядков возможно $r!$ и так как соединение всех получаемых таким образом областей D_r дает r -мерный куб $K_r [0 < x_i < t, i = 1, 2, \dots, r]$, то мы можем интегрировать по всему этому кубу с последующим делением на $r!$. Таким образом, мы получаем

$$P_k(t) = e^{-\lambda t} \sum_{r=k}^{\infty} \frac{\lambda^r}{r!} \sum_C \int_{K_r} \prod_{i=1}^k F(t - x_{s_i}) \prod_{\substack{s=1 \\ s \neq s_i}}^r [1 - F(t - x_s)] dx_1 \dots dx_r.$$

Здесь интеграл по кубу K_r распадается на произведение r простых интегралов и равен

$$\begin{aligned} \left\{ \int_0^t F(t-u) du \right\}^k \left\{ \int_0^t [1 - F(t-u)] du \right\}^{r-k} &= \\ &= \left\{ \int_0^t F(z) dz \right\}^k \left\{ \int_0^t [1 - F(z)] dz \right\}^{r-k}. \end{aligned}$$

Полагая

$$\int_0^t F(z) dz = v(t),$$

мы имеем

$$\int_0^t F(z) dz = 1 - \varepsilon(t), \quad \int_0^t [1 - F(z)] dz = t - 1 + \varepsilon(t),$$

и следовательно (так как число сочетаний C равно $\binom{r}{k}$),

$$\begin{aligned} P_k(t) &= e^{-\lambda t} \sum_{r=k}^{\infty} \frac{\lambda^r}{r!} \binom{r}{k} [1 - \varepsilon(t)]^k [t - 1 + \varepsilon(t)]^{r-k} = \\ &= e^{-\lambda t} \frac{\lambda^k}{k!} [1 - \varepsilon(t)]^k \sum_{r=k}^{\infty} \frac{\lambda^{r-k}}{(r-k)!} [t - 1 + \varepsilon(t)]^{r-k} = \\ &= e^{-\lambda t} \frac{\lambda^k}{k!} [1 - \varepsilon(t)]^k e^{\lambda[t-1+\varepsilon(t)]} = \\ &= e^{-\lambda} \frac{\lambda^k}{k!} [1 - \varepsilon(t)]^k e^{\lambda \varepsilon(t)}. \end{aligned}$$

Так как $\varepsilon(t) \rightarrow 0$ при $t \rightarrow \infty$, то отсюда

$$\lim_{t \rightarrow \infty} P_k(t) = e^{-\lambda} \frac{\lambda^k}{k!} \quad (k=0, 1, 2, \dots),$$

что и требовалось доказать *).

Глава 8

ЗАДАЧА ПАЛЬМА

§ 26. Постановка задачи

Для решения задачи Эрланга в гл. 6 и 7 нам было безразлично, рассматривается ли данный пучок линий как упорядоченный или нет. В настоящей главе, напротив, мы будем иметь дело с задачей, которая имеет смысл лишь для упорядоченного пучка. Мы будем, таким образом, все время исходить из предположения, что линии данного пучка перенумерованы и что каждый поступающий вызов занимает линию с наименьшим номером из числа тех, которые свободны в момент его поступления. Очевидно, что при таком порядке

*) В этом доказательстве мы неявно предполагали, что поток вызовов начинается с момента 0, т. е. что $P_0(0) = 1$, $P_k(0) = 0$ ($k > 0$). Доказательство лишь немного усложнилось бы при произвольных начальных данных.

средняя загруженность будет различной для различных линий: наиболее загруженной окажется первая линия, за ней вторая, и т. д.

С другой стороны, задачи, рассматриваемые в настоящей главе, таковы, что уже по смыслу их постановки безразлично, имеем ли мы дело с конечным или бесконечным пучком; таким образом, это различие, столь важное, как мы видели, для задачи Эрланга, для целей настоящей главы совершенно несущественно.

Во всем остальном мы сохраняем предпосылки предшествующих глав. Входящий поток вызовов мы предполагаем простейшим с параметром λ . Длительности разговоров предполагаются не зависящими ни друг от друга, ни от каких-либо данных о поступлении вызовов и распределенными по показательному закону со средним значением 1.

Условимся обозначать через L_r линию с номером r . Для всех рассуждений настоящей главы имеет основное значение тот простой и самоочевидный факт, что совокупность линий $L_1, L_2, \dots, L_r$ (при любом r , не превосходящем общего числа линий пучка) мы можем рассматривать как самостоятельный полнодоступный пучок. Каждый вызов, «потерянный» на этом пучке (т. е. заставший первые r линий занятыми), поступает на линию L_{r+1} (если, конечно, таковая имеется), и обратно — для того, чтобы вызов поступил на L_{r+1} , необходимо, чтобы он был потерян на пучке $(L_1, L_2, \dots, L_r)$. Вероятность потери на этом пучке есть доля времени, в течение которой все линии $L_1, L_2, \dots, L_r$ заняты; она, очевидно, совершенно не зависит от того, существуют ли еще линии с более высокими номерами и сколько таких линий; она может быть вычислена по формуле Эрланга для вероятности потери на пучке из r линий и равна

$$\frac{\frac{\lambda^r}{r!}}{\sum_{i=0}^r \frac{\lambda^i}{i!}}.$$

В частности, при $r=1$ мы получаем для вероятности потери вызова на линии L_1 выражение

$$\frac{\lambda}{1+\lambda}, \quad (26.1)$$

а при $r=2$ — для вероятности потери на пучке (L_1, L_2) выражение

$$\frac{\frac{\lambda^2}{2}}{1 + \lambda + \frac{\lambda^2}{2}} = \frac{\lambda^2}{\lambda^2 + 2\lambda + 2}. \quad (26.2)$$

Но потерю вызова на пучке (L_1, L_2) мы можем рассматривать как двойное событие: 1) потеря на L_1 и 2) потеря на L_2 . Вероятность первого из этих событий равна $\lambda/(1 + \lambda)$. Чтобы найти условную вероятность второго события при условии, что первое состоялось, заметим, что это есть вероятность потери на L_2 для вызова, потерянного на L_1 (или, что то же, поступившего на L_2). Если обозначить через λ' интенсивность потока вызовов, поступающих на L_2 , то искомая условная вероятность второго события будет поэтому по формуле (26.1) равна $\lambda'/(1 + \lambda')$ (так как для этого потока L_2 служит первой линией). Но λ' есть число вызовов, теряющихся на L_1 в единицу времени; так как поступает на L_1 в единицу времени λ вызовов, а доля потерь составляет $\lambda/(1 + \lambda)$, то

$$\lambda' = \lambda \frac{\lambda}{1 + \lambda} = \frac{\lambda^2}{1 + \lambda},$$

и искомая условная вероятность второго события равна

$$\frac{\lambda'}{1 + \lambda'} = \frac{\lambda^2}{1 + \lambda + \lambda^2}.$$

Для вероятности потери на пучке (L_1, L_2) мы таким образом получаем выражение

$$\frac{\lambda}{1 + \lambda} \frac{\lambda^2}{1 + \lambda + \lambda^2} = \frac{\lambda^3}{(1 + \lambda)(1 + \lambda + \lambda^2)}, \quad (26.3)$$

отличное от непосредственно даваемого формулами Эрланга выражения (26.2). Следовательно, формула (26.3) для вероятности потери на пучке (L_1, L_2) является ошибочной*). Анализируя цепь рассуждений, приведших нас к этой формуле, мы легко находим источник ошибки. Обозначив через λ' интенсивность потока вызовов, поступающих на L_2 , мы

*) Указанием на этот поучительный «парадокс» я обязан В. К. Лезерсону.

приняли, следуя формуле Эрланга (26.1), вероятность потери на L_2 для поступающего на эту линию вызова равной $\lambda'/(1+\lambda')$; тем самым мы неявно допустили, что поток вызовов, поступающих на L_2 , является *простейшим*, так как формулы Эрланга установлены нами лишь в этом предположении. То, что мы пришли к неверному результату, доказывает, что это предположение было ошибочным. Мы можем, таким образом, считать установленным, что *поток вызовов, поступающих на L_2 (или, что то же, теряемых на L_1), не есть простейший поток*. Тем более, конечно, у нас нет оснований ожидать, чтобы простейшими оказались потоки вызовов, поступающих на $L_3, L_4, \dots$. Этот отрицательного характера вывод поучителен тем, что ясно показывает, насколько важно уже для решения самых элементарных задач не ограничиваться изучением одних только простейших поступающих потоков.

Детальное, до конца идущее исследование природы потока вызовов, поступающих на любую линию L_r данного упорядоченного полnodоступного пучка, представляет собой георетически интересную и практически важную задачу, решению которой и будет посвящена настоящая глава. Все основные результаты в этом направлении были получены Пальмом [8] в 1943 г.

§ 27. Элементарные расчеты

Как мы уже указали в предыдущем параграфе, для вызова, поступающего на данный пучок (или, что то же, на линию L_1), вероятность оказаться потерянным на пучке ($L_1, L_2, \dots, L_r$) равна по формуле Эрланга

$$E_r = \frac{\frac{\lambda^r}{r!}}{\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}} \quad (r = 1, 2, \dots).$$

Очевидно, что вызов теряется на линии L_r тогда и только тогда, когда он теряется на пучке ($L_1, L_2, \dots, L_r$). Поэтому можно также сказать, что число E_r выражает собой вероятность потери на линии L_r ; однако при этом необходимо отчетливо иметь в виду, что речь идет о вероятности

потери на L_r для вызова, поступающего на L_1 ; вероятность же Π_r потери на L_r для вызова, поступающего на L_r , имеет другую величину, которую мы теперь должны постараться найти. Мы можем при этом допустить, что $r > 1$, так как, очевидно, $\Pi_1 = E_1$.

Вероятность E_r того, что вызов, поступивший на L_1 , будет потерян на L_r , очевидно, может быть представлена в виде произведения двух множителей: вероятности E_{r-1} того, что он будет потерян на L_{r-1} , и условной вероятности его потери на L_r , если известно, что он потерян на L_{r-1} (или, что то же, поступил на L_r). Но эта условная вероятность и есть Π_r ; поэтому

$$E_r = E_{r-1} \Pi_r,$$

или

$$\Pi_r = \frac{E_r}{E_{r-1}} \quad (r > 1).$$

Несмотря на свою кажущуюся простоту, эта формула для расчетов неудобна тем, что содержит одновременно E_r и E_{r-1} . Поэтому удобнее заменить ее формулой

$$\Pi_r = \frac{\lambda}{r + \lambda E_{r-1}} \quad (r > 1), \quad (27.1)$$

которую мы сейчас докажем. Положим для краткости

$$\sum_{k=0}^q \frac{\lambda^k}{k!} = S_q \quad (q \geq 0),$$

так что

$$E_r = \frac{1}{S_r} \frac{\lambda^r}{r!} \quad (r \geq 1);$$

следовательно, при $r > 1$

$$\frac{1}{\Pi_r} = \frac{E_{r-1}}{E_r} = \frac{r}{\lambda} \frac{S_r}{S_{r-1}} = \frac{r}{\lambda} \left[1 + \frac{\lambda^r / r!}{S_{r-1}} \right] = \frac{1}{\lambda} (r + \lambda E_{r-1}),$$

а это и есть (27.1). При этом необходимо иметь в виду, что λ в формуле (27.1) означает интенсивность первичного потока вызовов, поступающих на L_1 .

Представляет существенный интерес сравнить между собой вероятности потери на различных линиях при одинаковой интенсивности поступающих на них потоков. Произведенные

подсчеты во всех случаях показывают, что эта вероятность возрастает с номером линии. В цитированной нами работе Пальм утверждает, что это непосредственно вытекает из формулы (27.1). Мы не видим, однако, как можно было бы это показать. Более того, нам вообще неизвестно, верно ли это утверждение в его общей формулировке*). Нам удалось доказать в этом направлении только следующее значительно более скромное предложение.

Теорема. Вероятность потери на L_r при $r > 1$ всегда больше, чем вероятность потери на L_1 , если поступающие на эти линии потоки имеют одинаковую интенсивность.

Доказательство. Если интенсивность поступающего на L_1 потока равна λ , то вероятность поступить на L_r для вызова, поступившего на L_1 , равна E_{r-1} (так как это есть вероятность потерпеть потерю на линиях $L_1, L_2, \dots, L_{r-1}$). Поэтому среди λ вызовов, поступающих в среднем на L_1 в единицу времени, на L_r будет в среднем поступать λE_{r-1} вызовов, т. е. поток вызовов, поступающих на L_r , будет иметь интенсивность λE_{r-1} . При этом, как мы знаем, вероятность потери на L_r для вызовов этого потока равна

$$P_r = \frac{\lambda}{r + \lambda E_{r-1}} \quad (r > 1).$$

Если бы поток той же интенсивности λE_{r-1} падал на L_1 , то вероятность потери на L_1 для вызовов этого потока согласно формуле (26.1) была бы

$$\frac{\lambda E_{r-1}}{1 + \lambda E_{r-1}} \quad (r > 1).$$

Для доказательства нашей теоремы достаточно поэтому убедиться, что при $r > 1$ и $\lambda > 0$ мы всегда имеем

$$\frac{\lambda E_{r-1}}{1 + \lambda E_{r-1}} < \frac{\lambda}{r + \lambda E_{r-1}}, \quad \text{или} \quad r + \lambda E_{r-1} < \lambda + \frac{1}{E_{r-1}},$$

или, наконец,

$$\lambda E_{r-1} + r - \lambda < \frac{1}{E_{r-1}} \quad (r > 1). \quad (27.2)$$

*) Этот факт доказан в статье Т. А. Азларова «Обобщение одной теоремы А. Я. Хинчина», Труды Ташкент. гос. университета им. В. И. Ленина, вып. 189, 113—118, 1961.— Б. Г.

Пусть сначала $r=2$, $E_{r-1}=E_1=\frac{\lambda}{1+\lambda}$. Тогда (27.2) получает вид

$$\frac{\lambda^2}{1+\lambda} + 2 - \lambda < \frac{1+\lambda}{\lambda},$$

что равносильно

$$\frac{\lambda^2}{1+\lambda} + 1 - \lambda < \frac{1}{\lambda},$$

или

$$\frac{1}{1+\lambda} < \frac{1}{\lambda},$$

и, следовательно, выполняется при любом $\lambda > 0$.

Пусть теперь неравенство (27.2) справедливо при каком-нибудь $r > 1$ (и любом $\lambda > 0$); покажем, что в таком случае оно останется верным и при замене r на $r+1$, т. е. что при любом $\lambda > 0$ будет иметь место неравенство

$$\lambda E_r + r + 1 - \lambda < \frac{1}{E_r}.$$

В силу (27.1) [помня, что $\Pi_r = E_r/E_{r-1}$] мы имеем

$$E_r = \frac{\lambda E_{r-1}}{r + \lambda E_{r-1}}.$$

Отсюда легко находим

$$\begin{aligned} \lambda E_r + r + 1 - \lambda - \frac{1}{E_r} &= r \left\{ \frac{\lambda E_{r-1} + r - \lambda}{\lambda E_{r-1} + r} - \frac{1}{\lambda E_{r-1}} \right\} = \\ &= \frac{r H_r(\lambda)}{\lambda E_{r-1} (\lambda E_{r-1} + r)}, \end{aligned}$$

где

$$\begin{aligned} H_r(\lambda) &= \lambda E_{r-1} (\lambda E_{r-1} + r - \lambda) - (\lambda E_{r-1} + r) = \\ &= \lambda E_{r-1} \left(\lambda E_{r-1} + r - \lambda - \frac{1}{E_{r-1}} \right) - (\lambda E_{r-1} + r - \lambda) = \\ &= (\lambda E_{r-1} - 1) (\lambda E_{r-1} + r - \lambda) - \lambda. \end{aligned}$$

Наша теорема будет доказана, если мы убедимся, что $H_r(\lambda) < 0$ при любом $\lambda > 0$. Положим для краткости

$$\lambda E_{r-1} - 1 = A, \quad \lambda E_{r-1} + r - \lambda = B,$$

так что

$$H_r(\lambda) = AB - \lambda.$$

Мы можем допустить, что $AB > 0$, так как в противном случае тривиальным образом $H_r(\lambda) < 0$. Если $A > 0$ и $B > 0$, то, так как в силу (27.2) $B < 1/E_{r-1}$, мы имеем

$$AB < \frac{\lambda E_{r-1} - 1}{E_{r-1}} < \lambda;$$

если же $A < 0$, $B < 0$, то, очевидно,

$$|A| < 1, |B| = -B = \lambda - r - \lambda E_{r-1} < \lambda, AB = |A||B| < \lambda.$$

Таким образом, $H_r(\lambda) < 0$ во всех случаях, и наша теорема доказана.

§ 28. Основная теорема Пальма

Мы убедились в § 26, что если на линию L_1 поступает простейший поток вызовов, то поток, поступающий на линию L_r ($r > 1$), уже не будет простейшим. Важнейшая основная теорема теории Пальма состоит в том, что при этом на любую линию L_r *поступает поток типа Р, т. е. стационарный, ординарный и с ограниченным последствием*. От простейшего такой поток отличается, следовательно, только тем, что требование отсутствия последствия заменяется более общим требованием ограниченности последствия.

Для доказательства этой теоремы нам не понадобится никаких расчетов; достаточно лишь более внимательно взглянуть на картину происходящего. Так как для L_1 утверждение теоремы тривиально, то надо только показать, что, если оно верно для L_r , оно остается верным и для L_{r+1} ($r = 1, 2, \dots$); а так как вызовы, поступающие на L_{r+1} , совпадают с вызовами, теряющимися на L_r , то мы должны доказать следующее: *если на линию L_r поступает поток вызовов типа Р, то потерянные на L_r вызовы также образуют поток типа Р*. При такой постановке задачи, очевидно, самое существование линии L_{r+1} является несущественным.

Обозначим для краткости через A поток вызовов, поступающих на L_r , и через B — поток вызовов, теряющихся на L_r . Течение потока B после произвольно выбранного момента t_0 будет однозначно определено, если станет известно, сколько времени будет еще длиться занимающий линию L_r в момент t_0 разговор, а также каковы моменты поступающих после t_0 вызовов потока A и какова длительность начинаемых этими

вызовами разговоров. Но все эти факторы в свою очередь однозначно определяются течением потока A , который есть поток типа P и, следовательно, стационарен. Поэтому все перечисленные факторы не зависят от выбранного момента t_0 , а вместе с ними не зависит от t_0 и дальнейшее течение потока B ; другими словами, поток B также стационарен.

Ординарность потока B с самоочевидностью вытекает из того, что он составляет собой часть (ординарного) потока A .

Покажем, наконец, что поток B — с ограниченным последствием. Пусть $t_0 = \tau_0 = 0$; обозначим через t_i ($i = 1, 2, \dots$) моменты следующих за t_0 вызовов потока A и через $\tau_1, \tau_2, \dots$ — моменты следующих за t_0 вызовов потока B и положим

$$\tau_k - \tau_{k-1} = \zeta_k \quad (k = 1, 2, \dots).$$

Наша задача — показать, что закон распределения величины ζ_{k+1} не зависит от значений величин $\zeta_1, \zeta_2, \dots, \zeta_k$, или, что то же, от значений величин $\tau_1, \tau_2, \dots, \tau_k$. Но величина ζ_{k+1} будет однозначно определена, если станут известны:

- 1) расстояния от τ_k до поступающих на L_r после момента τ_k дальнейших вызовов;
- 2) остаточная длительность разговора, занимающего линию L_r в момент τ_k ;
- 3) длительности разговоров, занимающих линию L_r после момента τ_k .

Все эти три фактора независимы от значений, принимаемых величинами $\tau_1, \tau_2, \dots, \tau_k$. Для первого это следует из того, что поток A — с ограниченным последствием; для второго — из показательного закона распределения длин разговоров, а для третьего это самоочевидно. Но так как этими тремя факторами значение случайной величины ζ_{k+1} определяется однозначно, то и эта случайная величина не зависит от величин $\tau_1, \tau_2, \dots, \tau_k$, или, что то же, от величин $\zeta_1, \zeta_2, \dots, \zeta_k$. Но это и означает, что поток B — с ограниченным последствием. Таким образом, основная теорема Пальма доказана.

Отсюда, очевидно, следует, что для упорядоченного полностью доступного пучка с простейшим входящим потоком и показательным распределением длин разговоров поток вызовов, поступающих на любую линию L_r этого пучка, представляет собой поток типа P . Но такой поток (см. § 13) однозначно

определяется заданием функции Пальма $\varphi(t)$. Условимся в дальнейшем обозначать через $\varphi_r(t)$ ($r=0, 1, 2, \dots$) функцию Пальма для потока вызовов, поступающих на линию L_{r+1} (так что, в частности, $\varphi_0(t)=e^{-\lambda t}$). Мы видим, что наша задача сводится к определению функции $\varphi_r(t)$ для любого $r > 0$; этому и будет посвящено все дальнейшее.

§ 29. Вывод основной системы уравнений

Пусть в момент t_0 теряется вызов на линии L_r (или, что то же, поступает вызов на линию L_{r+1}). Тогда $\varphi_r(t)$ есть вероятность того, что в промежутке (t_0, t_0+t) ни один вызов не будет потерян на L_r (не поступит на L_{r+1}). Но это событие может произойти двумя способами:

(А) В промежутке (t_0, t_0+t) на L_r не поступит ни одного вызова.

(В) В промежутке (t_0, t_0+t) на L_r будут поступать вызовы, но ни один из них не будет потерян.

Мы имеем поэтому

$$\varphi_r(t) = P(A) + P(B);$$

при этом, в случае $r > 0$, $P(A) = \varphi_{r-1}(t)$, так как, с одной стороны, вызов, потерянный в момент t_0 на L_r , был потерян и на L_{r-1} , а с другой — поток вызовов, поступающих на L_r , совпадает с потоком вызовов, теряемых на L_{r-1} .

Переходим к определению $P(B)$. Пусть первый вызов, поступающий на L_r после момента t_0 , происходит в промежутке (t_0+x, t_0+x+dx) ; вероятность этого события равна $\varphi_{r-1}(x) - \varphi_{r-1}(x+dx) = -d\varphi_{r-1}(x)$. Для того чтобы этот вызов не был потерян на L_r , очевидно, необходимо и достаточно, чтобы линия L_r до момента t_0+x освободилась от того разговора, которым она была занята в момент t_0 ; вероятность этого события равна $1 - e^{-x}$. Таким образом, вероятность того, что первый после момента t_0 вызов поступит на L_r в промежутке (t_0+x, t_0+x+dx) и что этот вызов не будет потерян, равна

$$-(1 - e^{-x}) d\varphi_{r-1}(x).$$

Мы утверждаем теперь, что если наступило все описанное и если $x < t$, то вероятность того, что в остающемся промежутке (t_0+x, t_0+t) не будет потеряно на L_r ни

одного вызова, равна $\varphi_r(t-x)$. Это вытекало бы непосредственно из определения функции $\varphi_r(t)$, если бы вызов, поступивший на L_r в промежутке (t_0+x, t_0+x+dx) , был потерян на этой линии; но на самом деле этот вызов по нашему допущению не теряется, так что наше утверждение требует обоснования. Будет ли вызов, поступивший на L_r в момент t_0+x [точнее: в промежутке (t_0+x, t_0+x+dx)], потерян или нет, во всяком случае, раз этот вызов произошел, линия L_r с момента t_0+x будет занята. Когда она освободится — это в силу показательного закона распределения длин разговоров совершенно не зависит от того, была ли она занята поступившим в момент t_0+x вызовом или была занята ранее (и поступивший в момент t_0+x вызов был потерян). С другой стороны, моменты дальнейших (после t_0+x) поступающих на L_r вызовов зависят от того факта, что такой вызов поступил в момент t_0+x , но совершенно не зависят от судьбы этого вызова (от того, был ли он потерян или нет); не зависят, конечно, от этой судьбы и длины тех разговоров, которые начинаются этими последующими вызовами. Таким образом, ни один из факторов, определяющих собой наличие или отсутствие потерь на L_r в промежутке (t_0+x, t_0+t) , не зависит от того, какая судьба постигла вызов, поступивший в момент t_0+x . И хотя этот вызов по нашему предположению не был потерян, вероятность того, что в промежутке (t_0+x, t_0+t) потерь на линии L_r не будет, такова же, как если бы он был потерян, т. е. равна $\varphi_r(t-x)$. Сопоставляя это с тем, что было установлено ранее, мы приходим к следующему выводу: при $x < t$ вероятность того, что первый после момента t_0 вызов поступит на L_r в промежутке (t_0+x, t_0+x+dx) и что между t_0 и t_0+t ни один вызов не будет потерян на L_r , равна

$$-(1 - e^{-x}) d\varphi_{r-1}(x) \varphi_r(t-x).$$

Но чтобы получить вероятность события B , мы, очевидно, должны просуммировать все такие вероятности по x от 0 до t . Это дает

$$P(B) = - \int_0^t (1 - e^{-x}) \varphi_r(t-x) d\varphi_{r-1}(x),$$

и следовательно,

$$\varphi_r(t) = \varphi_{r-1}(t) - \int_0^t (1 - e^{-x}) \varphi_r(t-x) d\varphi_{r-1}(x) \\ (r \geq 1). \quad (29.1)$$

Это и есть исходная система уравнений теории Пальма. Из (29.1) непосредственно видно, что при любом $t > 0$

$$\varphi_r(t) \geq \varphi_{r-1}(t)$$

— неравенство, которое является очевидным и само по себе.

§ 30. Преобразование Лапласа

С целью решения основной системы уравнений (29.1) мы теперь заменим искомые функции $\varphi_r(t)$ их преобразованиями Лапласа. Положим

$$\psi_r(t) = \int_0^\infty e^{-tx} \varphi_r(x) dx \quad (r \geq 0).$$

Напишем уравнение (29.1) в виде

$$\varphi_r(x) = \varphi_{r-1}(x) - \int_0^x (1 - e^{-z}) \varphi_r(x-z) d\varphi_{r-1}(z),$$

умножим обе части на e^{-tx} и проинтегрируем по x от 0 до ∞ . Это дает

$$\begin{aligned} \psi_r(t) &= \psi_{r-1}(t) - \int_0^\infty e^{-tx} dx \int_0^x (1 - e^{-z}) \varphi_r(x-z) d\varphi_{r-1}(z) = \\ &= \psi_{r-1}(t) - \int_0^\infty (1 - e^{-z}) e^{-tz} d\varphi_{r-1}(z) \int_z^\infty e^{-t(x-z)} \varphi_r(x-z) dx = \\ &= \psi_{r-1}(t) - \int_0^\infty (1 - e^{-z}) e^{-zt} d\varphi_{r-1}(z) \int_0^\infty e^{-ty} \varphi_r(y) dy = \\ &= \psi_{r-1}(t) - \psi_r(t) \int_0^\infty (1 - e^{-z}) e^{-zt} d\varphi_{r-1}(z). \end{aligned} \quad (30.1)$$

Интеграция по частям легко дает

$$\int_0^{\infty} e^{-zt} d\varphi_{r-1}(z) = -1 + t\psi_{r-1}(t),$$

и, следовательно,

$$\int_0^{\infty} e^{-z(t+1)} d\varphi_{r-1}(z) = -1 + (t+1)\psi_{r-1}(t+1);$$

поэтому из (30.1) мы получаем

$$\psi_r(t) = \psi_{r-1}(t) - \psi_r(t) [t\psi_{r-1}(t) - (t+1)\psi_{r-1}(t+1)],$$

откуда

$$\psi_r(t) = \frac{\psi_{r-1}(t)}{1 + t\psi_{r-1}(t) - (t+1)\psi_{r-1}(t+1)}. \quad (30.2)$$

Таким образом, для определения функций $\psi_r(t)$ мы получаем простую рекуррентную формулу. Так как $\varphi_0(x) = e^{-\lambda x}$, то

$$\psi_0(t) = \int_0^{\infty} e^{-(\lambda+t)x} dx = \frac{1}{t+\lambda},$$

и соотношение (30.2) позволяет последовательно определить все функции $\psi_r(t)$. В частности, мы непосредственно видим, что все эти функции рациональны. Однако мы не можем удовлетвориться этим, так как для обратного перехода к функциям $\varphi_r(x)$ нам важно более детально знать свойства рациональных дробей $\psi_r(t)$. В частности, для этого обратного перехода существенное значение имеет разложение функций $\psi_r(t)$ на простые дроби и, следовательно, природа и расположение корней их знаменателей. Этими вопросами мы и должны будем теперь заняться.

Заметим еще, что простым преобразованием искомых функций

$$t\psi_r(t) = \chi_r(t) \quad (r=0, 1, 2, \dots)$$

мы можем привести рекуррентные формулы (30.2) к более краткому виду:

$$\chi_r(t) = \frac{\chi_{r-1}(t)}{1 + \chi_{r-1}(t) - \chi_{r-1}(t+1)};$$

впрочем, в дальнейшем мы этим замечанием пользоваться не будем.

§ 31. Определение функций $\psi_r(t)$

Обозначим через $B_r(t)$ ($r = -1, 0, 1, 2, \dots$) многочлен степени $r+1$:

$$B_r(t) = \lambda^{r+1} + \sum_{l=0}^r \binom{r+1}{l} t(t+1) \dots (t+r-l) \lambda^l,$$

где $\sum_{l=0}^{-1} = 0$, и следовательно, $B_{-1}(t) = \lambda^0 = 1$.

Лемма. Многочлены $B_r(t)$ связаны рекуррентной формулой

$$B_r(t) = tB_{r-1}(t+1) + \lambda B_{r-1}(t) \quad (r=0, 1, \dots) \quad (31.1)$$

Доказательство. Мы имеем

$$\lambda B_{r-1}(t) = \lambda^{r+1} + \sum_{l=0}^{r-1} \binom{r}{l} t(t+1) \dots (t+r-1-l) \lambda^{l+1};$$

$$\begin{aligned} tB_{r-1}(t+1) &= t\lambda^r + \sum_{l=0}^{r-1} \binom{r}{l} t(t+1) \dots (t+r-l) \lambda^l = \\ &= \sum_{l=0}^r \binom{r}{l} t(t+1) \dots (t+r-l) \lambda^l. \end{aligned}$$

Так как $\binom{r}{0} = \binom{r+1}{0}$ и, при $l > 0$, $\binom{r}{l-1} + \binom{r}{l} = \binom{r+1}{l}$, то отсюда

$$\begin{aligned} \lambda B_{r-1}(t) + tB_{r-1}(t+1) &= \\ &= \lambda^{r+1} + \sum_{l=1}^r \binom{r+1}{l} t(t+1) \dots (t+r-l) \lambda^l + \\ &\quad + \binom{r}{0} t(t+1) \dots (t+r) = \\ &= \lambda^{r+1} + \sum_{l=0}^r \binom{r+1}{l} t(t+1) \dots (t+r-l) \lambda^l = B_r(t), \end{aligned}$$

что и доказывает лемму.

Теперь мы можем найти явное выражение для рациональных функций $\psi_r(t)$.

Теорема.

$$\psi_r(t) = \frac{B_{r-1}(t+1)}{B_r(t)} \quad (t=0, 1, 2, \dots). \quad (31.2)$$

Доказательство. Так как $B_{-1}(t) = 1$ и, как легко видеть, $B_0(t) = t + \lambda$, то при $r=0$ доказываемое соотношение (31.2) имеет вид:

$$\psi_0(t) = \frac{1}{t + \lambda},$$

и было нами установлено уже в § 30. Допустим поэтому, что $r > 0$ и соотношение (31.2) уже установлено для $\psi_{r-1}(t)$. Убедимся, что оно остается справедливым и для $\psi_r(t)$; этим, очевидно, теорема будет доказана.

В силу принятого допущения мы имеем

$$\begin{aligned} 1 + t\psi_{r-1}(t) - (t+1)\psi_{r-1}(t+1) &= \\ &= 1 + \frac{tB_{r-2}(t+1)}{B_{r-1}(t)} - \frac{(t+1)B_{r-2}(t+2)}{B_{r-1}(t+1)} = \\ &= \frac{K_r(t)}{B_{r-1}(t)B_{r-1}(t+1)}, \end{aligned}$$

где

$$\begin{aligned} K_r(t) &= B_{r-1}(t+1)B_{r-1}(t) + tB_{r-2}(t+1)B_{r-1}(t+1) - \\ &\quad - (t+1)B_{r-2}(t+2)B_{r-1}(t). \end{aligned} \quad (31.3)$$

Поэтому в силу (30.2)

$$\begin{aligned} \psi_r(t) &= \frac{B_{r-2}(t+1)}{B_{r-1}(t)} \frac{B_{r-1}(t)B_{r-1}(t+1)}{K_r(t)} = \\ &= \frac{B_{r-2}(t+1)B_{r-1}(t+1)}{K_r(t)}. \end{aligned}$$

Подлежащее доказательству соотношение (31.2) поэтому равносильно соотношению

$$K_r(t) = B_r(t)B_{r-2}(t+1),$$

которое в свою очередь в силу (31.1) равносильно соотношению

$$K_r(t) = tB_{r-1}(t+1)B_{r-2}(t+1) + \lambda B_{r-1}(t)B_{r-2}(t+1);$$

подставляя вместо $K_r(t)$ его выражение (31.3), мы находим,

что доказать надо следующее равенство:

$$B_{r-1}(t+1)B_{r-1}(t) - (t+1)B_{r-2}(t+2)B_{r-1}(t) = \\ = \lambda B_{r-1}(t)B_{r-2}(t+1),$$

или

$$B_{r-1}(t+1) = (t+1)B_{r-2}(t+2) + \lambda B_{r-2}(t+1).$$

Но это соотношение мы непосредственно получаем, заменяя в рекуррентной формуле (31.1) r на $r-1$ и t на $t+1$. Таким образом, наша теорема доказана.

Мы видим, что каждая функция $\psi_r(t)$ представляет собой правильную рациональную дробь, числитель которой есть многочлен степени r , а знаменатель — степени $r+1$.

§ 32. Разложение функций $\psi_r(t)$ на простые дроби

Чтобы определить функции $\varphi_r(x)$, лапласовскими преобразованиями которых служат функции $\psi_r(t)$, мы должны теперь посмотреть, как рациональные функции $\psi_r(t)$ разлагаются на простые дроби, и с этой целью исследовать корни многочленов $B_r(t)$, служащих знаменателями этих рациональных функций.

Многочлен $B_r(t)$ степени $r+1$ имеет, как непосредственно видно из его определения, положительные коэффициенты, причем коэффициент при t^{r+1} равен 1. Отсюда уже следует, что все вещественные корни этого многочлена отрицательны и что он может быть представлен в виде:

$$B_r(t) = (t + a_{r,0})(t + a_{r,1}) \dots (t + a_{r,r}),$$

где числа $a_{r,i}$ ($0 \leq i \leq r$) положительны или мнимы. Мы утверждаем, что числа $a_{r,i}$ все положительны и что, если они расположены в порядке возрастания, то

$$a_{r,i} > a_{r,i-1} + 1 \quad (1 \leq i \leq r),$$

т. е. расстояние между двумя соседними корнями превосходит единицу.

Докажем это утверждение посредством индукции. Мы уже видели, что $B_0(t) = t + \lambda$, так что для $r=0$ наше утверждение верно. Допустим, что оно верно для $B_r(t)$ ($r \geq 0$), и покажем, что в таком случае оно справедливо и для

$B_{r+1}(t)$. Для этого мы воспользуемся рекуррентной формулой (31.1), в силу которой

$$B_{r+1}(t) = tB_r(t+1) + \lambda B_r(t) = \\ = t(t+1+a_{r0}) \dots (t+1+a_{rr}) + \lambda(t+a_{r0}) \dots (t+a_{rr});$$

при этом в силу нашего предположения все $a_{rk} > 0$ и $a_{rk} > a_{r,k-1} + 1$ ($1 \leq k \leq r$). Эта формула (как, впрочем, и непосредственное определение) показывает, что $B_{r+1}(0) > 0$. С другой стороны, она дает

$$B_{r+1}(-a_{r0}) = -a_{r0}(a_{r1} + 1 - a_{r0}) \dots (a_{rr} + 1 - a_{r0}) < 0.$$

Это показывает, что $B_{r+1}(t)$ имеет корень между 0 и $-a_{r0}$. Пусть теперь k — одно из чисел ряда 0, 1, ..., $r-1$. Тогда

$$B_{r+1}(-a_{rk}-1) = \lambda(a_{r0}-a_{rk}-1) \dots (a_{r,k-1}-a_{rk}-1) \times \\ \times (a_{rk}-a_{rk}-1)(a_{r,k+1}-a_{rk}-1) \dots (a_{rr}-a_{rk}-1)$$

имеет, как легко подсчитать, $k+1$ отрицательных множителей и, следовательно, знак $(-1)^{k+1}$; напротив

$$B_{r+1}(-a_{r,k+1}) = -a_{r,k+1}(-a_{r,k+1}+1+a_{r0}) \dots \\ \dots (-a_{r,k+1}+1+a_{rk})(-a_{r,k+1}+1+a_{r,k+1}) \dots \\ \dots (-a_{r,k+1}+1+a_{rr})$$

имеет $k+2$ отрицательных множителей и, следовательно, знак $(-1)^k$. Таким образом, многочлен $B_{r+1}(t)$ для любого k ($0 \leq k \leq r-1$) имеет корень в промежутке между $-a_{rk}-1$ и $-a_{r,k+1}$. Наконец, так как старший член многочлена $B_{r+1}(t)$ есть t^{r+2} , то при отрицательных t , достаточно больших по абсолютной величине, $B_{r+1}(t)$ имеет знак $(-1)^{r+2}$; в то же время в выражении

$$B_{r+1}(-a_{rr}-1) = \lambda(-a_{rr}-1+a_{r0}) \dots (-a_{rr}-1+a_{rr})$$

все скобки отрицательны, так что оно имеет знак $(-1)^{r+1}$. Это показывает, что $B_{r+1}(t)$ имеет корень между $-a_{rr}-1$ и $-\infty$.

Сопоставляя все полученное, мы видим, что многочлен $B_{r+1}(t)$ имеет по меньшей мере по одному корню в каждом из $r+2$ промежутков:

$$(0, -a_{r0}), (-a_{rk}-1, -a_{r,k+1}) (0 \leq k \leq r-1), \\ (-a_{rr}-1, -\infty),$$

которые попарно не имеют общих точек. Так как $B_{r+1}(t)$ есть многочлен степени $r+2$, то этим его корни исчерпаны; а так как рассмотренные нами $r+2$ промежутков таковы, что расстояние между двумя соседними из них равно единице, то расстояние между двумя соседними корнями многочлена $B_{r+1}(t)$ всегда превосходит единицу. Таким образом, наше утверждение о расположении корней многочленов $B_r(t)$ полностью доказано.

Из этого следует, что разложение функции $\psi_r(t)$ на простые дроби имеет вид:

$$\psi_r(t) = \sum_{k=0}^r \frac{C_{rk}}{t + a_{rk}},$$

где числители C_{rk} легко могут быть выражены через числа a_{rk} хорошо известными методами.

§ 33. Заключение

Так как функция $\frac{C_{rk}}{t + a_{rk}}$ служит, как легко непосредственно убедиться, преобразованием Лапласа функции $C_{rk}e^{-a_{rk}x}$, то функция $\psi(t)$, вид которой мы только что нашли, есть преобразование Лапласа функции

$$\sum_{k=0}^r C_{rk}e^{-a_{rk}x}. \quad (33.1)$$

Но мы определили $\psi_r(t)$ как преобразование Лапласа искомой функции $\varphi_r(x)$. Можем ли мы отсюда заключить, что $\varphi_r(x)$ совпадает с функцией (33.1)?

Теория преобразований Лапласа (в детали которой мы здесь не можем входить) показывает, что среди функций, обладающих данным преобразованием Лапласа, может быть только одна ограниченная и неотрицательная при $0 \leq x < +\infty$; а так как функция $\varphi_r(x)$, очевидно, обладает обоими этими свойствами, то для ее совпадения с функцией (33.1) достаточно убедиться в положительности этой последней; для этого же в свою очередь, очевидно, достаточно показать, что $C_{rk} \geq 0$ ($0 \leq k \leq r$). В своем цитированном нами исследовании Пальм дает явное выражение чисел C_{rk} через

числа a_{rk} и путем анализа этих выражений действительно доказывает положительность всех коэффициентов C_{rk} . Таким образом, мы имеем для всех $r \geq 0$

$$\varphi_r(x) = \sum_{k=0}^r C_{rk} e^{-a_{rk}x} \quad (0 \leq x < +\infty),$$

и задача Пальма может считаться полностью решенной. Мы видим, что для любого r функция Пальма $\varphi_r(x)$, однозначно определяющая собой поток вызовов, теряемых на L_r (т. е. поступающих на L_{r+1}), представляет собой линейную комбинацию $r+1$ показательных функций. В частности, Пальм приводит следующее явное выражение для $\varphi_1(x)$:

$$\begin{aligned} \varphi_1(x) = & \left(\frac{1}{2} + \frac{1}{4 \sqrt{\lambda + \frac{1}{4}}} \right) e^{-\left(\lambda + \frac{1}{2} - \sqrt{\lambda + \frac{1}{4}} \right) x} + \\ & + \left(\frac{1}{2} - \frac{1}{4 \sqrt{\lambda + \frac{1}{4}}} \right) e^{-\left(\lambda + \frac{1}{2} + \sqrt{\lambda + \frac{1}{4}} \right) x}. \end{aligned}$$

ЧАСТЬ III

СИСТЕМЫ С ОЖИДАНИЕМ

Глава 9

СЛУЧАЙ ПОКАЗАТЕЛЬНОГО РАСПРЕДЕЛЕНИЯ ДЛИН РАЗГОВОРОВ

§ 34. Вероятности состояний

Мы обращаемся теперь к изучению полноступного пучка с ожиданием. Поступившему вызову приходится ждать разговора тогда и только тогда, когда он застает все n линий пучка занятыми. Поэтому здесь наша прежняя «вероятность потери» (т. е. вероятность застать все линии занятыми) может быть названа «вероятностью ожидания». Эта величина понятным образом играет известную роль в оценке качества работы пучка. Однако для систем с ожиданием эта роль сравнительно невелика, так как, если даже значительному большинству вызовов приходится ожидать, обслуживание должно быть признано вполне удовлетворительным во всех тех случаях, когда промежутки ожидания оказываются в своем большинстве очень малыми. Решающую роль здесь играет не частота ожиданий («потерь»), а природа времени ожидания γ как случайной величины; частота же ожиданий дает нам только один штрих этой картины — вероятность неравенства $\gamma > 0$. Понятно поэтому, что конечной целью в исследовании систем с ожиданием всегда служит отыскание *закона распределения времени ожидания* γ .

Для всех задач, которые мы будем рассматривать, безразлично, является ли данный пучок линий упорядоченным или нет. Входящий поток вызовов мы будем всегда предполагать простейшим с параметром λ . Вызовы обслуживаются

в порядке их поступления. Длины разговоров всегда будут мыслиться независимыми как друг от друга, так и от течения потока вызовов. Что касается закона распределения этих длин, то именно он составляет собой основной момент различия в задачах теории систем с ожиданием. Обычно бывает так, что при различных распределениях длин разговоров к исследованию времени ожидания приходится подходить различными методами. В настоящей главе мы рассмотрим самый простой случай, когда длины l разговоров подчиняются показательному закону

$$P\{l > t\} = e^{-\beta t}, \quad t > 0, \quad \beta > 0 — \text{постоянная.}$$

Для этого случая полное решение задачи было дано еще Эрлангом [7].

Условимся обозначать через $P_k(t)$ вероятность того, что в момент t система находится в «состоянии k », т. е. имеется всего k «наличных» (говорящих или ожидающих) вызовов. При $k \leq n$ занято k линий, ожидающих нет; при $k > n$ все n линий заняты и имеется $k - n$ ожидающих.

Если $k < n$, то мы находимся в тех же условиях, что и в случае систем с потерями, так как при $k < n$ нет ни потерь, ни ожиданий. Все соображения, приведенные нами в § 20, поэтому остаются в силе и, как там, приводят нас к системе уравнений

$$P'_0(t) = -\lambda P_0(t) + \beta P_1(t);$$

$$P'_k(t) = \lambda P_{k-1}(t) - (\lambda + k\beta) P_k(t) + (k+1)\beta P_{k+1}(t) \quad (0 < k < n)$$

[ср. (6) § 20, где мы полагали $\beta = 1$]. Но последнее уравнение системы (6) должно быть теперь заменено другим, так как теперь возможен переход в состояние n из состояния $n+1$, которое не имело смысла в случае системы с потерями.

Проведем общее рассмотрение для любого $k \geq n$. Будем, как прежде, обозначать через $P_{rs}(\tau)$ вероятность перехода системы из состояния r в состояние s за время τ и понимать знак $\approx$ как равенство с точностью до бесконечно малых вида $o(\tau)$ при $\tau \rightarrow 0$. Тогда мы легко находим, аналогично § 20, при $k \geq n$ и $\tau \rightarrow 0$

$$P_{k-1,k}(\tau) \approx \lambda\tau, \quad P_{k+1,k}(\tau) \approx n\beta\tau; \\ P_{kk}(\tau) \approx 1 - \lambda\tau - n\beta\tau, \quad P_{ik}(\tau) \approx 0 \quad (|i - k| > 1)$$

[отличие от случая $k < n$ состоит в том, что теперь мы имеем $P_{k+1, k}(\tau) \approx n\beta\tau$, а не $\approx (k+1)\beta\tau$, как прежде, так как при $k \geq n$ в состоянии $k+1$ мы имеем n , а не $k+1$ занятых линий]. Эти оценки с помощью рассуждений, в точности аналогичных проведенным нами в § 20, приводят к соотношению

$$P_k(t+\tau) \approx P_{k-1}(t)\lambda\tau + P_k(t)(1-\lambda\tau - n\beta\tau) + P_{k+1}(t)n\beta\tau,$$

откуда мы, снова в тесной аналогии с рассуждениями § 20, получаем с помощью предельного перехода

$$P'_k(t) = \lambda P_{k-1}(t) - (\lambda + n\beta) P_k(t) + n\beta P_{k+1}(t) \quad (k \geq n).$$

Таким образом, в целом система (6') § 20 теперь заменяется (бесконечной) системой уравнений

$$\left. \begin{aligned} P'_0(t) &= -\lambda P_0(t) + \beta P_1(t); \\ P'_k(t) &= \lambda P_{k-1}(t) - (\lambda + k\beta) P_k(t) + (k+1)\beta P_{k+1}(t) \\ &\quad (0 < k < n); \\ P'_k(t) &= \lambda P_{k-1}(t) - (\lambda + n\beta) P_k(t) + n\beta P_{k+1}(t) \\ &\quad (k \geq n). \end{aligned} \right\} \quad (34.1)$$

Как и в случае систем с потерями, мы принимаем в качестве вероятностей состояний пределы, к которым стремятся вероятности $P_k(t)$ при $t \rightarrow \infty$. В случае систем с потерями мы доказали во всей полноте существование этих пределов [теорема Маркова, § 19]. В случае систем с ожиданием [т. е. в случае уравнений (34.1)] такое доказательство также может быть проведено; однако здесь оно несравненно более сложно и требует существенно новых идей, так как метод Маркова тесно связан с предположением конечного числа возможных состояний системы. Мы не можем поместить этого доказательства здесь и вынуждены ограничиться ссылкой на его возможность*). Необходимо еще отметить, что в нашем новом случае мы, кроме существования пределов величины $P_k(t)$ при $t \rightarrow \infty$, должны еще доказать возможность предельного

*) Мы не знаем в литературе ни одного изложения этого доказательства, хотя сама задача рассматривалась приводимым здесь методом много раз (Эрланг [7], Фрай [4], Колмогоров [2], Феллер [3]). Эрланг вводит возможность предельного перехода как особый постулат; все остальные упомянутые авторы ограничиваются кратким указанием на возможность доказательства.

перехода во всей системе (34.1) — вопрос, который в § 20 у нас не возникал, так как там мы имели дело с конечной системой уравнений.

Итак, мы допускаем, что при $t \rightarrow \infty$ существуют пределы

$$\lim_{t \rightarrow \infty} P_k(t) = p_k \quad (k=0, 1, \dots)$$

и что соответствующий предельный переход возможен одновременно во всех уравнениях системы (34.1). Так как при этом левые части всех этих уравнений в пределе обращаются в нуль (это доказывается в точности так же, как и в § 20), то мы приходим к системе линейных уравнений

$$\left. \begin{aligned} -\lambda p_0 + \beta p_1 &= 0; \\ \lambda p_{k-1} - (\lambda + k\beta) p_k + (k+1)\beta p_{k+1} &= 0 \quad (0 < k < n); \\ \lambda p_{k-1} - (\lambda + n\beta) p_k + n\beta p_{k+1} &= 0 \quad (k \geq n), \end{aligned} \right\} \quad (34.2)$$

из которой, в соединении с нормирующим условием

$$\sum_{k=0}^{\infty} p_k = 1$$

мы и должны определить числа p_k .

Полагая

$$\lambda p_{k-1} - k\beta p_k = z_k \quad (1 \leq k \leq n),$$

мы легко находим из системы (34.2)

$$z_1 = 0, \quad z_k - z_{k+1} = 0 \quad (1 \leq k \leq n),$$

откуда

$$z_k = 0 \quad (0 \leq k \leq n),$$

или

$$p_k = \frac{\lambda}{k\beta} p_{k-1} \quad (0 < k \leq n).$$

Это дает, если для краткости положить еще $\lambda/\beta = y$,

$$p_k = \frac{y^k}{k!} p_0 \quad (0 \leq k \leq n);$$

в частности, при $k=n$ мы находим

$$p_n = \frac{y^n}{n!} p_0. \quad (34.3)$$

Чтобы найти значения p_k при $k > n$, мы обращаемся к последней группе уравнений (34.2). Перепишем их в виде:

$$n\beta(p_{k+1} - p_k) = \lambda(p_k - p_{k-1}) \quad (k \geq n),$$

и просуммируем по k от n до $n+r$:

$$n\beta(p_{n+r+1} - p_n) = \lambda(p_{n+r} - p_{n-1}),$$

откуда

$$n\beta p_{n+r+1} + z_n = \lambda p_{n+r},$$

или, так как $z_n = 0$, $\lambda/\beta = y$,

$$p_{n+r+1} = \frac{y}{n} p_{n+r} \quad (r \geq 0),$$

и следовательно, в силу (34.3)

$$p_{n+r+1} = \left(\frac{y}{n}\right)^{r+1} \frac{y^n}{n!} p_0 \quad (r \geq 0).$$

Таким образом, для любого $k \geq n$ мы имеем

$$p_k = \left(\frac{y}{n}\right)^{k-n} p_n = \frac{y^k}{n! n^{k-n}} p_0. \quad (34.4)$$

Соединяя этот результат с полученным прежде для $k \leq n$, мы находим

$$\left. \begin{aligned} p_k &= \frac{y^k}{k!} p_0 & (0 \leq k \leq n); \\ p_k &= \frac{y^k}{n! n^{k-n}} p_0 & (k \geq n). \end{aligned} \right\} \quad (34.5)$$

Нам остается найти p_0 . Нормирующее условие

$$\sum_{k=0}^{\infty} p_k = 1$$

дает

$$\frac{1}{p_0} = \sum_{k=0}^{n-1} \frac{y^k}{k!} + \frac{y^n}{n!} \sum_{k=n}^{\infty} \left(\frac{y}{n}\right)^k = s_{n-1}(y) + \frac{y^n}{(n-1)!(n-y)},$$

где положено $s_m(y) = \sum_{k=0}^m (y^k/k!)$. Вспомним для сравнения, что в случае системы с потерями мы имели (§ 20)

$$\frac{1}{p_0} = s_n(y) = s_{n-1}(y) + \frac{y^n}{n!}.$$

Отметим еще, что вероятность найти все линии занятыми («вероятность ожидания») равна в силу (34.4)

$$\Pi = \sum_{k=n}^{\infty} p_k = \frac{n^n p_0}{n!} \sum_{k=n}^{\infty} \left(\frac{y}{n}\right)^k = \frac{y^n}{n!} \frac{p_0}{1 - \frac{y}{n}}. \quad (34.6)$$

§ 35. Закон распределения времени ожидания

Теперь мы уже легко можем найти вероятность $P\{\gamma > t\}$ того, что для поступившего в произвольно выбранный момент вызова время ожидания будет больше чем t . Обозначим через $P_k\{\gamma > t\}$ условную вероятность того же неравенства в предположении, что произведенный вызов застал систему в состоянии k . По формуле полной вероятности мы имеем

$$P\{\gamma > t\} = \sum_{k=0}^{\infty} p_k P_k\{\gamma > t\},$$

или, так как, очевидно, $P_k\{\gamma > t\} = 0$ при $k < n$ и $t \geq 0$,

$$P\{\gamma > t\} = \sum_{k=n}^{\infty} p_k P_k\{\gamma > t\}. \quad (35.1)$$

Величины p_k нам известны; остается определить величины $P_k\{\gamma > t\}$ при всех $k \geq n$.

Положим $k - n = v$ ($v = 0, 1, 2, \dots$). Наша задача — найти вероятность неравенства $\gamma > t$ при условии, что в момент вызова все линии были заняты и сверх того имелось v ожидающих. Очевидно, что при этом наш вызов получает разговор после $(v + 1)$ -го освобождения линии. Искомая вероятность есть поэтому вероятность того, что за время t после появления нашего вызова произойдет не более чем v освобождений линии. Пусть $q_r(t)$ ($0 \leq r \leq v$) есть вероятность того, что за это время произойдет ровно r освобождений;

тогда в силу $k - n = v$

$$P_k\{\gamma > t\} = \sum_{r=0}^{k-n} q_r(t) \quad (k \geq n).$$

Но поток освобождений за время ожидания нашего вызова представляет собой в силу показательного закона распределения длин разговоров простейший поток с параметром $n\beta$, так как вероятность того, что не произойдет ни одного освобождения за время t после такого момента, когда все линии заняты, равна $(e^{-\beta t})^n = e^{-n\beta t}$. Величина $q_r(t)$ есть вероятность того, что за время t наступит r событий этого потока; в силу формул главы 1 поэтому

$$q_r(t) = e^{-n\beta t} \frac{(n\beta t)^r}{r!} \quad (0 \leq r \leq v),$$

и мы находим

$$P_k\{\gamma > t\} = \sum_{r=0}^{k-n} e^{-n\beta t} \frac{(n\beta t)^r}{r!} \quad (k \geq n).$$

Возвращаясь к формуле (35.1) и используя соотношение (34.4), мы находим

$$\begin{aligned} P\{\gamma > t\} &= \sum_{k=n}^{\infty} p_k \sum_{r=0}^{k-n} e^{-n\beta t} \frac{(n\beta t)^r}{r!} = \\ &= e^{-n\beta t} \sum_{k=n}^{\infty} \left(\frac{y}{n}\right)^{k-n} p_n \sum_{r=0}^{k-n} \frac{(n\beta t)^r}{r!} = \\ &= p_n e^{-n\beta t} \sum_{r=0}^{\infty} \frac{(n\beta t)^r}{r!} \sum_{k=n+r}^{\infty} \left(\frac{y}{n}\right)^{k-n} = \\ &= p_n e^{-n\beta t} \sum_{r=0}^{\infty} \frac{(n\beta t)^r}{n^r r!} \sum_{k=n+r}^{\infty} \left(\frac{y}{n}\right)^{k-n-r} = \\ &= \frac{p_n e^{-n\beta t}}{1 - \frac{y}{n}} \sum_{r=0}^{\infty} \frac{(\lambda t)^r}{r!} = \frac{p_n}{1 - \frac{y}{n}} e^{-(n\beta - \lambda)t}, \end{aligned}$$

или, так как в силу (34.3) и (34.6) $p_n = \frac{y^n}{n!} p_0 = \Pi \left(1 - \frac{y}{n}\right)$

$$P\{\gamma > t\} = \Pi e^{-(n\beta - \lambda)t} \quad (t \geq 0).$$

Этим наша задача решена. Мы видим, что в принятых нами условиях время ожидания подчиняется показательному закону распределения с параметром $\lambda\beta - \lambda$. Вместе с тем мы получаем

$$P\{\gamma > 0\} = \Pi,$$

как оно и должно быть: через Π мы как раз обозначили в конце § 34 вероятность заставить все линии занятыми («вероятность ожидания»).

Глава 10

ОДНОЛИНЕЙНЫЕ СИСТЕМЫ В СЛУЧАЕ СТАНДАРТНОЙ ДЛИНЫ РАЗГОВОРА

§ 36. Разностно-дифференциальное уравнение задачи

За пределами показательного распределения длин разговоров исследование систем с ожиданием сопряжено с большими трудностями. Простые и законченные результаты здесь удастся получить лишь в некоторых частных предположениях. Особенно важным в практическом отношении является случай систем с одной линией (короче: однолинейных систем), для которых задача исследования времени ожидания может быть продвинута весьма далеко при статистических предпосылках широкой общности. Этим случаем мы и будем теперь заниматься до конца книги.

В настоящей главе мы изложим созданную Эрлангом интересную как оригинальностью метода, так и законченностью результатов теорию однолинейных систем в предположении, что все разговоры имеют в точности одну и ту же длину τ . Во всех других отношениях мы сохраняем статистические предпосылки предшествующей главы.

Рассмотрим какие-либо два последовательных вызова. Пусть γ_0 — время ожидания первого вызова, а γ — второго. Обозначим через z расстояние между этими двумя вызовами. Очевидно, что если $z \geq \gamma_0 + \tau$, то при появлении второго вызова (единственная) линия свободна и $\gamma = 0$; если же $z < \gamma_0 + \tau$, то при появлении второго вызова линия занята и ему приходится ждать в течение промежутка времени $\gamma = \gamma_0 + \tau - z$. Таким образом, мы находим, что при данном

$$z = u$$

$$\gamma = \begin{cases} 0 & (\gamma_0 + \tau - u \leq 0); \\ \gamma_0 + \tau - u & (\gamma_0 + \tau - u \geq 0), \end{cases}$$

или, что то же,

$$\gamma = \max(0, \gamma_0 + \tau - u).$$

Пусть теперь t — любое положительное число. Убедимся, что при данном $z = u$ неравенства $\gamma < t$ и $\gamma_0 < t + u - \tau$ равносильны между собой. В самом деле, если $\gamma < t$, то либо $\gamma = 0$, и тогда $\gamma_0 \leq u - \tau < u + t - \tau$, либо $\gamma = \gamma_0 + \tau - u$, и тогда $\gamma_0 + \tau - u < t$, и значит, $\gamma_0 < t + u - \tau$; обратно, если $\gamma_0 < t + u - \tau$, то либо $\gamma_0 < u - \tau$, $\gamma_0 + \tau - u < 0$, $\gamma = 0 < t$, либо $\gamma_0 > u - \tau$, $\gamma = \gamma_0 + \tau - u < t$. Таким образом, если P_u означает условную вероятность, вычисленную в предположении, что $z = u$, то мы имеем при любом $u > 0$

$$P_u \{\gamma < t\} = P_u \{\gamma_0 < t + u - \tau\} \quad (t > 0);$$

но γ_0 (время ожидания первого вызова), очевидно, не зависит (как случайная величина) от того, когда последует второй вызов, т. е. какое значение получит величина z ; поэтому условная вероятность $P_u \{\gamma_0 < t + u - \tau\}$ неравенства $\gamma_0 < t + u - \tau$ равна безусловной вероятности $P \{\gamma_0 < t + u - \tau\}$ того же неравенства, и мы получаем

$$P_u \{\gamma < t\} = P \{\gamma_0 < t + u - \tau\} \quad (t > 0). \quad (36.1)$$

Обозначим через $f(t)$ закон распределения величины γ , полагая

$$f(t) = P \{\gamma < t\}.$$

Так как поток вызовов мы предполагаем простейшим с параметром λ , то вероятность неравенств $u < z < u + du$ (с точностью до бесконечно малых высших порядков) равна $\lambda e^{-\lambda u} du$, и мы по формуле полной вероятности находим

$$f(t) = \int_0^{\infty} \lambda e^{-\lambda u} P_u \{\gamma < t\} du,$$

или в силу (36.1) при $t \geq 0$

$$f(t) = \int_0^{\infty} \lambda e^{-\lambda u} P \{\gamma_0 < t + u - \tau\} du.$$

Но закон распределения величины γ_0 дается той же функцией $f(t)$, что и для величины γ ; поэтому мы находим

$$f(t) = \int_0^{\infty} \lambda e^{-\lambda u} f(t+u-\tau) du \quad (t \geq 0).$$

Это уравнение и будет служить нам основой для определения искомой функции $f(t)$. Прежде всего, предполагая эту функцию дифференцируемой, мы находим

$$f'(t) = \int_0^{\infty} \lambda e^{-\lambda u} f'(t+u-\tau) du \quad (t \geq 0),$$

и интеграция по частям дает

$$\begin{aligned} f'(t) &= \lambda e^{-\lambda u} f(t+u-\tau) \Big|_0^{\infty} + \lambda^2 \int_0^{\infty} e^{-\lambda u} f(t+u-\tau) du = \\ &= -\lambda f(t-\tau) + \lambda f(t), \end{aligned}$$

или

$$f'(t) = \lambda [f(t) - f(t-\tau)] \quad (t \geq 0).$$

Мы получаем, таким образом, для определения функции $f(t)$ разностно-дифференциальное уравнение простого вида. Это уравнение может быть еще упрощено преобразованием искомой функции

$$f(t) = e^{\lambda t} g(t),$$

что, как легко видеть, дает для новой неизвестной функции уравнение

$$g'(t) = -\lambda e^{-\lambda \tau} g(t-\tau) \quad (t > 0). \quad (36.2)$$

§ 37. Закон распределения времени ожидания

Рассмотрим сначала отрезок времени $0 < t \leq \tau$; в силу $t > 0$ при этом имеют место все соотношения, выведенные в § 36. Но при $t \leq \tau$ мы имеем $f(t-\tau) = P\{\gamma < t-\tau\} = 0$, и следовательно, соотношение (36.2) дает

$$g'(t) = 0, \quad g(t) = c = \text{const} \quad (0 < t \leq \tau),$$

откуда

$$f(t) = c e^{\lambda t} \quad (0 < t \leq \tau).$$

Чтобы определить постоянную c , заставим в последнем равенстве t стремиться к нулю; в пределе мы получим $c = f(+0)$, но $f(-0) = 0$, а потому $c = f(+0) - f(-0) = P\{\gamma = 0\}$; это есть вероятность того, что поступившему в произвольно выбранный момент вызову не придется ожидать (линия окажется свободной); поэтому $c = 1 - \alpha$, где α означает вероятность в произвольно выбранный момент застать линию занятой. Иначе говоря, α есть математическое ожидание суммарной длительности всех разговоров, ведущихся в течение единицы времени. Но математическое ожидание числа разговоров в единицу времени равно λ , а длительность каждого разговора равна τ ; следовательно,

$$\alpha = \lambda\tau, \quad c = 1 - \alpha = 1 - \lambda\tau.$$

Таким образом, окончательно

$$f(t) = (1 - \alpha)e^{\lambda t} \quad (\alpha = \lambda\tau, \quad 0 < t \leq \tau),$$

и закон распределения времени ожидания найден нами для отрезка $0 < t \leq \tau$.

Мы теперь докажем, что для любого неотрицательного целого числа n в отрезке $n\tau < t \leq (n+1)\tau$ функция $g(t)$ определяется формулой

$$g(t) = (1 - \alpha) \sum_{k=0}^n e^{-\lambda k\tau} \frac{(k\tau - t)^k}{k!} \lambda^k. \quad (37.1)$$

Так как функция $f(t)$ просто выражается через функцию $g(t)$, то этим мы получим явное выражение для закона распределения времени ожидания.

При $n=0$ формула (37.1) дает

$$g(t) = 1 - \alpha \quad (0 < t \leq \tau),$$

что нами доказано выше. Поэтому мы можем доказать формулу (37.1) с помощью индукции. Допустим, что она верна для какого-либо числа $n \geq 0$, и покажем, что в таком случае она останется верной и для числа $n+1$.

Итак, пусть соотношение (37.1) верно при $n\tau < t \leq (n+1)\tau$; если $(n+1)\tau < t \leq (n+2)\tau$, то в силу (36.2)

мы будем иметь

$$\begin{aligned} g'(t) &= -\lambda e^{-\lambda \tau} g(t - \tau) = \\ &= -\lambda e^{-\lambda \tau} (1 - \alpha) \sum_{k=0}^n e^{-\lambda k \tau} \frac{(k\tau - t + \tau)^k}{k!} \lambda^k; \end{aligned}$$

интегрируя это равенство по t от $(n+1)\tau$ до числа, которое мы снова обозначим через t , найдем

$$\begin{aligned} g(t) - g[(n+1)\tau] &= \\ &= -\lambda e^{-\lambda \tau} (1 - \alpha) \sum_{k=0}^n \frac{e^{-\lambda k \tau} \lambda^k}{k!} \int_{(n+1)\tau}^t [(k+1)\tau - u]^k du. \end{aligned}$$

Здесь

$$\begin{aligned} \int_{(n+1)\tau}^t [(k+1)\tau - u]^k du &= \\ &= -\frac{1}{k+1} \{[(k+1)\tau - t]^{k+1} - [(k+1)\tau - (n+1)\tau]^{k+1}\}, \end{aligned}$$

а потому

$$\begin{aligned} g(t) - g[(n+1)\tau] &= \\ &= \lambda e^{-\lambda \tau} (1 - \alpha) \sum_{k=0}^n \frac{e^{-\lambda k \tau} \lambda^k}{(k+1)!} \{[(k+1)\tau - t]^{k+1} - \\ &\quad - [(k+1)\tau - (n+1)\tau]^{k+1}\} = \\ &= (1 - \alpha) \sum_{k=0}^n \frac{e^{-\lambda(k+1)\tau} \lambda^{k+1}}{(k+1)!} \{[(k+1)\tau - t]^{k+1} - \\ &\quad - [(k+1)\tau - (n+1)\tau]^{k+1}\} = \\ &= (1 - \alpha) \sum_{r=1}^{n+1} \frac{e^{-\lambda r \tau} \lambda^r}{r!} \{[r\tau - t]^r - [r\tau - (n+1)\tau]^r\} = \\ &= (1 - \alpha) \sum_{r=0}^{n+1} \frac{\lambda^r e^{-\lambda r \tau}}{r!} (r\tau - t)^r - \\ &\quad - (1 - \alpha) \sum_{r=0}^{n+1} \frac{\lambda^r e^{-\lambda r \tau}}{r!} [r\tau - (n+1)\tau]^r, \quad (37.2) \end{aligned}$$

где суммирование по r можно вести от нуля, так как члены с $r=0$ в обеих суммах, очевидно, взаимно уничтожаются.

Формула (37.1) по нашему допущению верна при $n\tau < t \leq (n+1)\tau$. Полагая в ней $t = (n+1)\tau$, мы находим

$$g[(n+1)\tau] = (1-\alpha) \sum_{r=0}^{n+1} \frac{\lambda^r e^{-\lambda r\tau}}{r!} [r\tau - (n+1)\tau]^r, \quad (37.3)$$

где суммирование можно вести до $n+1$ потому, что член суммы с $r=n+1$, очевидно, равен нулю. Наконец, складывая между собой равенства (37.2) и (37.3), мы находим

$$g(t) = (1-\alpha) \sum_{r=0}^{n+1} \frac{\lambda^r e^{-\lambda r\tau}}{r!} (r\tau - t)^r$$

при любом t в промежутке $(n+1)\tau < t \leq (n+2)\tau$, что и требовалось доказать.

Таким образом, мы получаем для любого $t > 0$

$$P\{\gamma < t\} = f(t) = e^{-\lambda t} (1 - \lambda\tau) \sum_{k=0}^n e^{-\lambda k\tau} \frac{(k\tau - t)^k}{k!} \lambda^k,$$

где целое неотрицательное число n определяется неравенствами

$$n\tau < t \leq (n+1)\tau.$$

Глава 11

ОБЩАЯ ТЕОРИЯ ОДНОЛИНЕЙНЫХ СИСТЕМ

§ 38. Постановка задачи и обозначения

В двух предшествующих главах мы, следуя классическим работам Эрланга, нашли закон распределения времени ожидания в двух наиболее простых предположениях относительно закона распределения длин разговоров: в случае показательного распределения (см. гл. 9) и (для однолинейных систем) в случае стандартной длительности разговора (см. гл. 10). Однако можно считать, что на практике мы встречаемся

с тем или другим из этих двух простейших распределений лишь в очень редких случаях. В большинстве же приложений мы можем в лучшем случае рассчитывать лишь на некоторое приближение реального распределения длин разговоров к той или другой из наших двух предпосылок; и даже для такого расчета в очень многих случаях мы не имеем в сущности никаких оснований. Поэтому представляется весьма желательным построить метод, позволяющий определить закон распределения времени ожидания (или хотя бы важнейшие его статистические характеристики) при возможно широких предпосылках относительно распределения длин разговоров.

В своей общей постановке эта задача приводит к расчетам, трудно обозримым по своей сложности. Поэтому во всем дальнейшем мы сосредоточим наше внимание на практически весьма важном случае системы с одной линией *). Зато в отношении закона распределения длин разговоров мы ограничимся естественным требованием существования конечного математического ожидания, оставляя этот закон во всем остальном совершенно произвольным. Мы увидим, что при этих предпосылках задача отыскания закона распределения времени ожидания принципиально решается до конца сравнительно несложными приемами.

В течение всей настоящей главы мы будем иметь дело с пучком из одной линии, на которую поступает простейший поток вызовов с параметром λ . Вызовы, заставшие линию занятой, ожидают ее освобождения и занимают ее в порядке их поступления. Длины разговоров не зависят ни друг от друга, ни от числа ожидающих. Вероятность того, что длина произвольно выбранного разговора окажется больше чем t , мы будем обозначать через $F(t)$, так что средняя длитель-

*) В последние годы появились многочисленные работы, посвященные однолинейным системам, в которых, с одной стороны, были получены интересные новые общие результаты, а с другой, намечены другие подходы к решению возникающих вопросов. Мы здесь ограничимся указанием лишь на небольшое число статей:

D. J. Kendall. Some problems in the theory of queues Journ. Royal Stat. Soc., Ser. B 1951 151—173.
 D. V. Lindley. The theory of queues with a single server, Proc. Cambr. Phil. Soc., v. 48, 1952, 277—289.
 L. Takács. Investigation of waiting time problems by reduction to Markov processes, Acta Math. Acad. Sci. Hungar., v. 6, 1955, 101—129 — Б. Г.

ность разговора будет

$$-\int_0^{\infty} t dF(t) = \int_0^{\infty} F(t) dt = s.$$

Кроме того, мы вводим следующие обозначения, которыми будем пользоваться на протяжении всей главы:

α — вероятность того, что в произвольно выбранный момент времени линия окажется занятой; иначе говоря математическое ожидание суммарного времени занятости линии за 1 час (часом мы будем условно называть принятую единицу времени);

π_k ($k=0, 1, 2, \dots$) — вероятность того, что в начале произвольно выбранного разговора мы будем иметь k ожидающих;

$v_k(t) = e^{-\lambda t} (\lambda t)^k / k!$ — вероятность того, что в течение промежутка времени длины t на линию поступит k вызовов;

γ (случайная величина) — время ожидания для вызова, поступившего в произвольно выбранный момент времени.

Дальнейшие обозначения будут объяснены по мере их введения.

§ 39. Вспомогательные предложения

Лемма 1. Пусть $0 < a < b$. Вероятность застать линию в произвольно выбранный момент времени занятой разговором, длина которого заключена между a и b , равна

$$-\lambda \int_a^b u dF(u).$$

Доказательство. Обозначим через $P_T(a, b)$ вероятность застать линию занятой разговором длины, заключенной между a и b^*), в момент времени, произвольно выбранный в промежутке $(0, T)$. Лемма 1 утверждает тогда, что

$$\lim_{T \rightarrow \infty} P_T(a, b) = -\lambda \int_a^b u dF(u).$$

Обозначим далее через $L_T^*(a, b)$ суммарную длительность всех разговоров и частей разговоров длины (a, b) , веду-

*) В дальнейшем мы такой разговор для краткости будем называть «разговором длины (a, b) ».

щихся в промежутке $(0, T)$. $L_T^*(a, b)$ есть случайная величина; если она принимает какое-либо значение l , то соответствующая условная вероятность заставить разговор длины (a, b) , очевидно, равна l/T . Поэтому в силу формулы полной вероятности

$$P_T(a, b) = \sum_l P \{L_T^*(a, b) = l\} \frac{l}{T} = \frac{1}{T} M L_T^*(a, b),$$

где M — символ математического ожидания.

Для доказательства леммы 1 достаточно поэтому установить, что при $T \rightarrow \infty$

$$\lim \frac{1}{T} M L_T^*(a, b) = -\lambda \int_a^b u dF(u). \quad (39.1)$$

Пусть $L_T(a, b)$ означает суммарную длительность разговоров длины (a, b) , начинающихся в отрезке $(0, T)$. Так как, очевидно,

$$|L_T(a, b) - L_T^*(a, b)| < b,$$

а следовательно, и

$$|M L_T(a, b) - M L_T^*(a, b)| < b,$$

то (39.1) равносильно соотношению

$$\frac{1}{T} M L_T(a, b) \rightarrow -\lambda \int_a^b u dF(u) \quad (T \rightarrow \infty);$$

а так как $M L_T(a, b)$, очевидно, пропорционально T , то последнее соотношение означает просто, что при любом $T > 0$

$$\frac{1}{T} M L_T(a, b) = -\lambda \int_a^b u dF(u).$$

В частности, это соотношение будет установлено (и, значит, лемма 1 доказана), если мы убедимся, что

$$\lim_{T \rightarrow \infty} \frac{1}{T} M L_T(a, b) = -\lambda \int_a^b u dF(u). \quad (39.2)$$

Пусть u — любое положительное число и $T < u$. В промежутке $(0, T)$ может тогда начаться не более одного разговора длины $(u, u + du)$, так что случайная величина $L_T(u, u + du)$ может, кроме значения 0, принять только значение вида $u + \theta du$, где $0 < \theta < 1$; такое значение она примет, очевидно, если (1°) в отрезке $(0, T)$ начнется по меньшей мере один разговор и (2°) последний из начавшихся в $(0, T)$ разговоров будет длины $(u, u + du)$. Вероятность (2°) есть $F(u) - F(u + du)$, а вероятность (1°) при $T \rightarrow 0$ имеет вид $\lambda T + o(T)$, где λ — среднее число начал разговоров в единицу времени, совпадающее, очевидно, с параметром входящего потока вызовов. Таким образом, случайная величина $L_T(u, u + du)$ в рассматриваемом нами случае либо равна нулю, либо принимает значение вида $u + \theta du$, причем последнее — с вероятностью $[\lambda T + o(T)][F(u) - F(u + du)]$. Таким образом,

$$ML_T(u, u + du) = (u + \theta du) [\lambda T + o(T)] [F(u) - F(u + du)],$$

и следовательно, суммируя по u от a до b ,

$$ML_T(a, b) = [-\lambda T + o(T)] \int_a^b u dF(u).$$

Отсюда следует (39.2), а значит, и лемма 1.

Лемма 1, таким образом, доказана. При $a = 0$, $b = +\infty$ рассматриваемая в ней вероятность, очевидно, есть вероятность α того, что в произвольно выбранный момент времени линия окажется занятой. Мы приходим, таким образом, к важной формуле

$$\alpha = -\lambda \int_0^{\infty} t dF(t) = \lambda s$$

(которую, впрочем, можно было бы усмотреть и непосредственным рассуждением).

Условимся называть *разговором типа k* ($k = 0, 1, 2, \dots$) разговор, в начале которого имеется k ожидающих, так что введенную нами в § 38 величину π_k можно определить как долю разговоров типа k среди всех ведущихся разговоров (вероятность того, что наудачу выбранный разговор окажется разговором типа k).

Лемма 2. Если в некоторый момент линия оказалась занятой, то условная вероятность того, что она занята разговором типа k , равна π_k .

Доказательство. По определению величины π_k среди λ разговоров, ведущихся в среднем в единицу времени, мы будем иметь в среднем $\lambda\pi_k$ разговоров типа k . Так как длительность разговора, как случайная величина, не зависит от числа ожидающих в его начале (т. е. от его типа), то средняя длительность разговора типа k равна s , а следовательно, суммарная длительность разговоров типа k , ведущихся в единицу времени, в среднем равна $\lambda\pi_k s = \alpha\pi_k$; но условная вероятность, о которой идет речь в лемме 2, есть отношение этой средней суммарной длительности ко времени занятости линии, т. е. к α ; таким образом, эта условная вероятность равна π_k , и лемма 2 доказана.

Выберем теперь произвольный разговор длины $t > 0$ и найдем вероятность $u_k(z)$ ($0 < z \leq t$) того, что через промежуток времени z после начала этого разговора число ожидающих будет равно k ($k=0, 1, 2, \dots$). Так как длина и тип разговора взаимно независимы, то вероятность того, что в начале выбранного разговора будет r ожидающих, равна π_r ; если же это случится, то для того, чтобы по истечении времени z число ожидающих достигло k , необходимо и достаточно, чтобы было $r \leq k$ и чтобы за этот промежуток времени z поступило $k-r$ новых вызовов, вероятность чего есть

$$v_{k-r}(z) = e^{-\lambda z} \frac{(\lambda z)^{k-r}}{(k-r)!}.$$

Мы приходим таким образом к следующему предположению.

Лемма 3. Вероятность $u_k(z)$ того, что через промежуток времени $z \leq t$ после начала произвольно выбранного разговора длины t мы будем иметь k ожидающих, равна

$$u_k(z) = \sum_{r=0}^k \pi_r v_{k-r}(z) \quad (k=0, 1, 2, \dots).$$

Условимся теперь обозначать через τ промежуток времени между произвольно выбранным моментом и окончанием

того разговора, который мы в этот момент застали (остаточная длительность случайно застигнутого разговора). Имеет место

Лемма 4. Пусть $k \geq 0$, $x > 0$ и $dx > 0$ мало. Вероятность застать в произвольно выбранный момент k ожидающих и разговор с $x < \tau < x + dx$ равна

$$-\lambda dx \int_x^{\infty} u_k(t-x) dF(t) + o(dx).$$

Доказательство. Трудность обоснования леммы 4 заключается в том, что закон распределения случайной величины τ зависит от того, сколько ожидающих мы застаем в выбранный нами момент времени. Если мы обозначим через a — событие, состоящее в том, что в выбранный момент имеется k ожидающих;

b — событие, состоящее в том, что в выбранный момент ведется разговор с $x < \tau < x + dx$,

то эти события взаимно зависимы; наша задача — определить вероятность $P(ab)$ их совместного наступления. С этой целью рассмотрим сначала условную вероятность $P_t(ab)$ той же пары событий при условии, что застигнутый разговор имеет длину t . Мы можем записать в легко понятных обозначениях

$$P_t(ab) = P_t(b) P_{tb}(a). \quad (39.3)$$

Так как выбранный нами момент, если он застал разговор длины t , с одинаковой вероятностью мог попасть в любой момент этого разговора, то

$$P_t(b) = \begin{cases} dx/t & (t > x); \\ 0 & (t \leq x). \end{cases} \quad (39.4)$$

С другой стороны событие b равносильно событию

$$t - x - dx < t - \tau < t - x,$$

где $t - \tau$ есть возраст застигнутого разговора. Таким образом, $P_{tb}(a)$ есть условная вероятность застать в произвольно выбранный момент k ожидающих, если известно, что в этот момент мы застали разговор длины t и возраста, заключенного между $t - x - dx$ и $t - x$. Согласно определению

функции $u_k(x)$ (см. лемму 3) эта вероятность при $dx \rightarrow 0$ асимптотически равна $u_k(t-x)^*$.

Таким образом, в силу (39.3) и (39.4) мы имеем

$$P_t(ab) = \begin{cases} u_k(t-x) dx/t + o(dx) & (t > x); \\ 0 & (t \leq x). \end{cases}$$

Такова условная вероятность совместного наступления событий a и b при условии, что застигнут разговор длины t . Нам остается теперь освободиться от этого условия. Согласно лемме 1 закон распределения длины застигнутого разговора есть

$$G(t) = -\lambda \int_t^{\infty} u \, dF(u),$$

и мы имеем по формуле полной вероятности

$$\begin{aligned} P(a, b) &= - \int_0^{\infty} P_t(a, b) \, dG(t) = \\ &= - \int_x^{\infty} u_k(t-x) \frac{dx}{t} \, dG(t) + o(dx) = \\ &= -\lambda \int_x^{\infty} u_k(t-x) \frac{dx}{t} t \, dF(t) + o(dx) = \\ &= -\lambda dx \int_x^{\infty} u_k(t-x) \, dF(t) + o(dx), \end{aligned}$$

что и требовалось доказать.

Лемма 5.

$$\begin{aligned} - \int_0^{\infty} u_{k+1}(x) \, dF(x) &= \pi_k \quad (k > 0); \\ - \int_0^{\infty} u_1(x) \, dF(x) &= \pi_0 \left\{ 1 + \int_0^{\infty} e^{-\lambda x} \, dF(x) \right\}. \end{aligned}$$

*) Непрерывность функции $u_k(z)$ вытекает из леммы 3.

Доказательство. По смыслу функции $u_{k+1}(x)$ [см. формулировку леммы 3] интеграл

$$-\int_0^{\infty} u_{k+1}(x) dF(x)$$

означает вероятность того, что в конце произвольно выбранного разговора мы будем иметь $k+1$ ожидающих. В случае $k > 0$ это равносильно тому, что следующий разговор начнется при k ожидающих, вероятность чего равна π_k . Этим доказано первое из двух утверждаемых равенств.

В случае $k=0$ вероятность того, что выбранный разговор окончится при одном ожидающем, будет *меньше* чем π_0 ; дело в том, что разговор без ожидающих может начаться не только при окончании предшествующего разговора с одним ожидающим, но и другим способом — появлением вызова в момент, когда линия свободна. Поэтому

$$-\int_0^{\infty} u_1(x) dF(x) = \pi_0 - q,$$

где $q > 0$. Для определения q мы сложим между собой все установленные нами соотношения, замечая при этом, что

$$\sum_{k=0}^{\infty} \pi_k = 1 \quad \text{и} \quad \sum_{k=0}^{\infty} u_k(x) = 1$$

при любом x ; поэтому мы получаем

$$-\int_0^{\infty} [1 - u_0(x)] dF(x) = 1 - q;$$

а так как в силу леммы 3 $u_0(x) = \pi_0 v_0(x) = \pi_0 e^{-\lambda x}$, то отсюда

$$q = -\pi_0 \int_0^{\infty} e^{-\lambda x} dF(x), \quad (39.5)$$

и лемма 5 доказана полностью.

Лемма 6. При $|a| \leq 1$, $z > 0$

$$\sum_{k=0}^{\infty} a^k u_k(z) = e^{\lambda z(a-1)} \sum_{k=0}^{\infty} \pi_k a^k.$$

Доказательство. В силу леммы 3

$$\begin{aligned}\sum_{k=0}^{\infty} a^k u_k(z) &= \sum_{k=0}^{\infty} a^k \sum_{r=0}^k \pi_r e^{-\lambda z} \frac{(\lambda z)^{k-r}}{(k-r)!} = \\ &= e^{-\lambda z} \sum_{r=0}^{\infty} \pi_r \sum_{k=r}^{\infty} a^k \frac{(\lambda z)^{k-r}}{(k-r)!} = e^{-\lambda z} \sum_{r=0}^{\infty} \pi_r a^r \sum_{k=r}^{\infty} \frac{(a\lambda z)^{k-r}}{(k-r)!} = \\ &= e^{\lambda z(a-1)} \sum_{k=0}^{\infty} \pi_k a^k,\end{aligned}$$

что и требовалось доказать.

Обозначим через $\psi(\xi)$ характеристическую функцию длительности разговора, т. е. положим

$$\psi(\xi) = - \int_0^{\infty} e^{i\xi t} dF(t);$$

положим, далее, для любого вещественного ξ

$$\chi(\xi) = \sum_{k=0}^{\infty} \pi_k [\psi(\xi)]^k;$$

наконец, пусть

$$\eta = \eta(\xi) = \frac{\lambda}{i} [\psi(\xi) - 1].$$

Лемма 7.

$$\chi(\xi) = (1 - \alpha) \frac{1 - \psi(\xi)}{\psi(\eta) - \psi(\xi)}$$

Доказательство. В силу леммы 5 мы имеем [определяя q формулой (39.5)]

$$\begin{aligned}\chi(\xi) &= \sum_{k=0}^{\infty} [\psi(\xi)]^k \pi_k = - \sum_{k=0}^{\infty} [\psi(\xi)]^k \int_0^{\infty} u_{k+1}(x) dF(x) + q = \\ &= q - \frac{1}{\psi(\xi)} \int_0^{\infty} \left\{ \sum_{k=1}^{\infty} [\psi(\xi)]^k u_k(x) \right\} dF(x),\end{aligned}$$

откуда в силу леммы 6

$$\begin{aligned}
 \chi(\xi) &= q - \frac{1}{\psi(\xi)} \int_0^{\infty} \left\{ e^{i\lambda\psi(\xi)-1} \sum_{k=0}^{\infty} \pi_k [\psi(\xi)]^k - u_0(x) \right\} dF(x) = \\
 &= q - \frac{1}{\psi(\xi)} \int_0^{\infty} \{ e^{\lambda\psi(\xi)-1} \chi(\xi) - \pi_0 e^{-\lambda x} \} dF(x) = \\
 &= q - \frac{\chi(\xi)}{\psi(\xi)} \int_0^{\infty} e^{\lambda\psi(\xi)-1} dF(x) + \frac{1}{\psi(\xi)} \int_0^{\infty} \pi_0 e^{-\lambda x} dF(x) = \\
 &= q - \frac{\chi(\xi)}{\psi(\xi)} \int_0^{\infty} e^{i\eta x} dF(x) - \frac{q}{\psi(\xi)} = q \frac{\psi(\xi)-1}{\psi(\xi)} + \frac{\chi(\xi)\psi(\eta)}{\psi(\xi)}; \\
 \chi(\xi) \frac{\psi(\xi)-\psi(\eta)}{\psi(\xi)} &= q \frac{\psi(\xi)-1}{\psi(\xi)}; \\
 \chi(\xi) &= q \frac{1-\psi(\xi)}{\psi(\eta)-\psi(\xi)}.
 \end{aligned}$$

Для доказательства леммы 7 остается показать, что $q = 1 - \alpha$. Но q не зависит от ξ , а так как $\chi(0) = 1$, то из последней формулы

$$q = \lim_{\xi \rightarrow 0} \frac{\psi(\eta) - \psi(\xi)}{1 - \psi(\xi)},$$

или, по правилу Лопиталья,

$$q = \lim_{\xi \rightarrow 0} \frac{\psi'(\eta) \frac{d\eta}{d\xi} - \psi'(\xi)}{-\psi'(\xi)} = 1 - \lim_{\xi \rightarrow 0} \frac{\psi'(\eta) \frac{d\eta}{d\xi}}{\psi'(\xi)};$$

но $\psi'(0) = i\lambda$, $\frac{d\eta}{d\xi} = \frac{\lambda}{i} \psi'(\xi) \rightarrow \lambda s = \alpha$ при $\xi \rightarrow 0$, и мы действительно находим $q = 1 - \alpha$.

§ 40. Характеристическая функция времени ожидания

Целью нашего исследования является отыскание закона распределения случайной величины γ — времени ожидания для вызова, производимого в произвольно выбранный момент времени. Так как всякий закон распределения однозначно

определяется соответствующей характеристической функцией, а отыскание важнейших свободных характеристик случайной величины по ее характеристической функции обычно бывает проще, чем по ее закону распределения, то мы естественно будем считать нашу цель достигнутой, если нам удастся найти характеристическую функцию $\varphi(\xi)$ величины γ , т. е. математическое ожидание величины $e^{i\gamma\xi}$ как функцию вещественного параметра ξ . К этому и будут направлены наши усилия.

Пусть в дальнейшем M есть символ математического ожидания, так что $\varphi(\xi) = M e^{i\gamma\xi}$. Мы будем различать следующие возможности. Во-первых, в произвольно выбранный момент мы можем застать линию свободной; вероятность этого равна $1 - \alpha$, и в этом случае наверняка $\gamma = 0$, $e^{i\gamma\xi} = 1$. Во-вторых, мы можем застать линию занятой при k ожидающих ($k = 0, 1, 2, \dots$); вероятность этого мы обозначим через $P(k)$, а математические ожидания, вычисленные при этом условии, будем обозначать через M_k . Очевидно, мы имеем

$$\varphi(\xi) = M e^{i\gamma\xi} = 1 - \alpha + \sum_{k=0}^{\infty} P(k) M_k e^{i\gamma\xi}.$$

Если наш вызов застал линию занятой при k ожидающих, то время ожидания γ для него складывается из двух частей: 1) остаточная длительность τ того разговора, который ведется в момент появления нашего вызова, и 2) суммарная длительность T разговоров тех k ожидающих, которых наш вызов застал при своем появлении. Мы имеем поэтому

$$\gamma = \tau + T, \quad M_k e^{i\gamma\xi} = M_k (e^{i\tau\xi} e^{iT\xi}).$$

Но очевидно, что при данном k величины τ и T взаимно независимы. Поэтому мы имеем

$$\varphi(\xi) = 1 - \alpha + \sum_{k=0}^{\infty} P(k) M_k e^{i\tau\xi} M_k e^{iT\xi}.$$

Величина $M_k e^{i\tau\xi}$ вычисляется непосредственно. В самом деле, величина T есть сумма k взаимно независимых случайных величин, распределенных по закону $F(x)$ с характеристической функцией $\psi(\xi)$. Поэтому

$$M_k e^{iT\xi} = [\psi(\xi)]^k \quad (k = 0, 1, 2, \dots)$$

и мы находим

$$\varphi(\xi) = 1 - \alpha + \sum_{k=0}^{\infty} [\psi(\xi)]^k P(k) M_k e^{i\xi\tau}.$$

Обозначим через $Q_k(x)$ закон распределения величины τ (вероятность неравенства $\tau > x$) при k ожидающих. Тогда

$$M_k e^{i\xi\tau} = - \int_0^{\infty} e^{i\xi x} dQ_k(x);$$

$$P(k) M_k e^{i\xi\tau} = - \int_0^{\infty} e^{i\xi x} P(k) dQ_k(x).$$

Но величина

$$P(k) [Q_k(x) - Q_k(x + dx)]$$

есть, очевидно, вероятность того, что наш вызов застанет линию занятой при k ожидающих и что при этом будет $x < \tau \leq x + dx$. Эта вероятность в силу леммы 4 с точностью до малых высших порядков при $dx \rightarrow 0$ равна

$$-\lambda dx \int_x^{\infty} u_k(t - x) dF(t).$$

Поэтому мы получаем

$$P(k) M_k e^{i\xi\tau} = -\lambda \int_0^{\infty} e^{i\xi x} dx \int_x^{\infty} u_k(t - x) dF(t),$$

и следовательно,

$$\varphi(\xi) = 1 - \alpha - \lambda \int_0^{\infty} e^{i\xi x} dx \int_x^{\infty} \left\{ \sum_{k=0}^{\infty} [\psi(\xi)]^k u_k(t - x) \right\} dF(t),$$

откуда в силу леммы 6

$$\begin{aligned} \varphi(\xi) &= 1 - \alpha - \lambda \int_0^{\infty} e^{i\xi x} dx \int_x^{\infty} e^{\lambda(t-x)[\psi(\xi)-1]} \sum_{k=0}^{\infty} \pi_k [\psi(\xi)]^k dF(t) = \\ &= 1 - \alpha - \lambda \chi(\xi) \int_0^{\infty} e^{i\xi x} dx \int_x^{\infty} e^{\lambda(t-x)[\psi(\xi)-1]} dF(t), \end{aligned}$$

где положено, как в § 39, $\sum_{k=0}^{\infty} \pi_k [\psi(\xi)]^k = \chi(\xi)$. Если мы еще в согласии с § 39 положим

$$\frac{\lambda}{i} [\psi(\xi) - 1] = \eta = \eta(\xi),$$

то получим

$$\begin{aligned} \varphi(\xi) &= 1 - \alpha - \lambda \chi(\xi) \int_0^{\infty} e^{i(\xi - \eta)x} dx \int_x^{\infty} e^{i\eta t} dF(t) = \\ &= 1 - \alpha - \lambda \chi(\xi) \int_0^{\infty} e^{i\eta t} dF(t) \int_0^t e^{i(\xi - \eta)x} dx = \\ &= 1 - \alpha - \lambda \chi(\xi) \int_0^{\infty} e^{i\eta t} dF(t) \frac{e^{i(\xi - \eta)t} - 1}{i(\xi - \eta)} = \\ &= 1 - \alpha - \frac{\lambda \chi(\xi)}{i(\xi - \eta)} \int_0^{\infty} [e^{i\xi t} - e^{i\eta t}] dF(t) = \\ &= 1 - \alpha + \frac{\lambda}{i} \chi(\xi) \frac{\psi(\xi) - \psi(\eta)}{\xi - \eta}. \end{aligned}$$

Наконец, выражая $\chi(\xi)$ согласно лемме 7, получаем

$$\begin{aligned} \varphi(\xi) &= 1 - \alpha + (1 - \alpha) \frac{\lambda}{i} \frac{\psi(\xi) - 1}{\xi - \eta} = \\ &= (1 - \alpha) \left\{ 1 + \frac{\lambda}{i} \frac{\psi(\xi) - 1}{\xi - \frac{\lambda}{i} [\psi(\xi) - 1]} \right\} = \\ &= (1 - \alpha) \frac{\frac{i\xi}{\lambda}}{\frac{\xi}{\lambda} - [\psi(\xi) - 1]} = \frac{1 - \alpha}{1 - \alpha \frac{\psi(\xi) - 1}{is\xi}}. \end{aligned}$$

Полученная таким образом формула

$$\varphi(\xi) = \frac{1 - \alpha}{1 - \alpha \frac{\psi(\xi) - 1}{is\xi}} \quad (40.1)$$

может считаться полным решением поставленной задачи, так как характеристическая функция $\varphi(\xi)$ искомого закона рас-

предела времени ожидания нами выражена через данные постоянные s и $\alpha = \lambda s$ и через характеристическую функцию $\varphi(\xi)$ данного закона распределения $F(t)$ длин разговоров.

В качестве первой иллюстрации формулы (40.1) рассмотрим, какое выражение с ее помощью получает среднее значение $\bar{\gamma}$ времени ожидания. Мы имеем в виду $\varphi(0) = 1$

$$i\bar{\gamma} = \varphi'(0) = \frac{\varphi'(0)}{\varphi(0)};$$

а так как в силу (40.1)

$$\varphi(\xi) = \frac{(1 - \alpha) is\xi}{is\xi - \alpha[\varphi(\xi) - 1]},$$

то

$$\begin{aligned} \frac{\varphi'(\xi)}{\varphi(\xi)} &= \frac{d \ln \varphi(\xi)}{d\xi} = \frac{1}{\xi} - \frac{is - \alpha\varphi'(\xi)}{is\xi - \alpha[\varphi(\xi) - 1]} = \\ &= \frac{\alpha\{\xi\varphi'(\xi) - [\varphi(\xi) - 1]\}}{\xi\{is\xi - \alpha[\varphi(\xi) - 1]\}}. \end{aligned}$$

Предел этого выражения при $\xi \rightarrow 0$ легко находится по правилу Лопиталья и равен

$$\frac{\alpha\psi''(0)}{2(1 - \alpha)is}.$$

Величина

$$\psi''(0) = \int_0^{\infty} t^2 dF(t) = -s_2$$

есть взятое с обратным знаком математическое ожидание квадрата длины разговора. Мы находим таким образом

$$i\bar{\gamma} = -\frac{\alpha}{2i(1 - \alpha)} \frac{s_2}{s},$$

или

$$\bar{\gamma} = \frac{\alpha}{2(1 - \alpha)} \frac{s_2}{s}.$$

Эта простая формула, между прочим, показывает, что при данной загрузке линии α и при данной средней длине разговора s время ожидания будет в среднем тем меньше, чем меньше дисперсия длин разговоров, т. е. чем более стандартизованы эти длины.

В качестве второй иллюстрации рассмотрим простейший случай показательного распределения

$$F(t) = e^{-\beta t}$$

длин разговоров. Мы легко находим в этом случае

$$s = \frac{1}{\beta}, \quad \psi(\xi) = \frac{\beta}{\beta - i\xi} = \frac{1}{1 - i\xi s},$$

откуда

$$\frac{\psi(\xi) - 1}{is\xi} = \frac{1}{1 - i\xi s} = \psi(\xi),$$

и, значит, в силу (40.1)

$$\varphi(\xi) = (1 - \alpha) \frac{1 - i\xi s}{1 - \alpha - i\xi s} = 1 - \alpha + \alpha \frac{1 - \alpha}{1 - \alpha - i\xi s}. \quad (40.2)$$

Но

$$\frac{1 - \alpha}{1 - \alpha - i\xi s} = \frac{\frac{1 - \alpha}{s}}{\frac{1 - \alpha}{s} - i\xi}$$

есть характеристическая функция показательного закона распределения с параметром $(1 - \alpha)/s$. Поэтому (40.2) показывает, что закон распределения времени ожидания есть

$$P\{\gamma \geq t\} = (1 - \alpha) E(t) + \alpha e^{-\frac{1 - \alpha}{s} t} \quad (t \geq 0),$$

где

$$E(t) = \begin{cases} 1 & (t \leq 0); \\ 0 & (t > 0). \end{cases}$$

Это согласуется с формулой

$$P\{\gamma > t\} = \Pi e^{-(n\beta - \lambda)t},$$

полученной нами в конце главы 10 для случая n линий. В самом деле, при $n=1$ мы имеем $\Pi = \alpha = \lambda s$, $n\beta - \lambda = \frac{1}{s} - \lambda = \frac{1 - \alpha}{s}$. Член же $(1 - \alpha) E(t)$ при $t > 0$ равен нулю, а при $t=0$ обращается в $1 - \alpha = P\{\gamma = 0\}$.

ЛИТЕРАТУРНЫЕ УКАЗАНИЯ

(числа в прямых скобках относятся к списку литературы в конце настоящих указаний на стр. 148).

- § 1. Эти элементарные сведения помещаются почти во всех современных курсах теорин вероятностей; см., в частности, Гнеденко [1], Феллер [3], Фрай [4], Хинчин [5], Erlang [7].
- § 2. По-видимому, публикуется впервые.
- § 3, 4. См. § 1.
- § 5. См. Фрай [4].
- § 6. В этой форме публикуется впервые.
- § 7. Публикуется впервые.
- § 8. См. Redheffer [9], где в другой форме решается та же задача.
- § 9, 10. Относительно функции $\varphi_0(t)$ см. Palm [8]. Относительно функций $\varphi_k(t)$ ($k > 0$) публикуется впервые.
- § 11. Публикуется впервые.
- § 12. Публикуется впервые.
- § 13. Понятие потока с ограниченным последствием принадлежит Пальму [8]. В остальном содержание параграфа публикуется впервые.
- § 14. См. Palm [8].
- § 15, 16. Публикуется впервые.
- § 17, 18, 19, 20. См. Феллер [3], Фрай [4], Erlang [7].
- § 21. Эта эргодическая теорема является частным случаем известных общих теорем эргодической теорин. В такой редакции, по-видимому, публикуется впервые.
- § 22, 23. Erlang [7].
- § 24. Публикуется впервые. Частные случаи см. Palm [8].
- § 25. Публикуется впервые.
- § 26. См. Palm [8].
- § 27. Публикуется впервые.

§ 28, 29, 30, 31, 32, 33. См. Palm [8].

§ 34, 35. См. Колмогоров [2], Феллер [3], Фрай [4], Erlang [7].

§ 36, 37. См. Erlang [7].

§ 38, 39, 40. Вся глава II представляет собой существенную переработку статьи автора [6].

ЛИТЕРАТУРА

1. Гнеденко Б. В., Курс теории вероятностей, изд. 2-е, Гостехиздат, 1954.
2. Колмогоров А. Н., Sur le problème d'attente, Mat. сборн., 1931, 38, № 1—2, 101—106.
3. Феллер В., Введение в теорию вероятностей и ее приложения, Изд. иностр. лит., 1952.
4. Фрай Т., Теория вероятностей для инженеров, Гостехиздат, 1934.
5. Хинчин А. Я., Асимптотические законы теории вероятностей, О.ИИ, 1936.
6. Хинчин А. Я., Математическая теория стационарной очереди, Mat. сборн. 1932, 39, № 4, 73—81.
7. Erlang A. K. The Life and Works of The Copenhagen Telephone Company, 1943.
8. Palm C., Intensitätsschwankungen im Fernspreverkehr, Ericson Technics, 1943, 44, 1—189.
9. Redheffer R. M., A note on the Poisson law, Math. Magazine, 1953, 26, № 4, 185—188.

МАТЕМАТИЧЕСКАЯ ТЕОРИЯ СТАЦИОНАРНОЙ ОЧЕРЕДИ*)

Предлагаемое исследование было предпринято с целью ответить на некоторые вопросы, возникшие в связи со строительством и эксплуатацией телефонных станций; однако, результаты его имеют совершенно общий характер и могут найти применение к случаю любой очереди в периоде ее стационарности, т. е. при условии, что очередь не имеет тенденции к нарастанию или убыванию; следует отметить, что применяемые здесь методы, безусловно, могут служить орудием исследования и нестационарных очередей; однако, эта (зрочем, практически менее важная) часть работы в настоящее время еще не выполнена.

В частности, я полагаю, что результаты настоящего исследования могут служить основанием для расчетов основных показателей времени простоя станков и других установок в некоторых отраслях промышленности — всякий раз, когда большее число установок поручается наблюдению одного рабочего. В этом случае очередь образуется станками, ожидающими ремонтной операции. Однако в случае, когда число станков, поручаемых одному рабочему, не велико (лежит в пределах одного десятка), формулы настоящей статьи становятся неточными, и следует избегать пользоваться ими. Этот случай, более сложный, требует специального исследования, которое я надеюсь изложить в другой работе.

Опытная проверка результатов настоящего исследования ведется Московской городской телефонной станцией. Результаты ее, а также практические приемы расчета времени

*) Матем. сб., т. 39 (1932), № 4, стр. 73—84.

ожидания, вытекающие из полученных формул, будут опубликованы в соответствующей научно-технической печати.

Некоторые из результатов настоящей работы содержатся в качестве частных случаев в работах Pollaczek'a 1930 г. *) Однако я счел желательным все же их опубликовать, так как чрезвычайно громоздкий метод (позволяющий, правда, в то же время решать более общие задачи) делает чтение его работ крайне затруднительным, в особенности для инженера. Некоторые частные случаи формулы для среднего времени ожидания (§ 4) были найдены Thornton Fry'em**), которому, по-видимому, принадлежит и идея применяемого мною рекуррентного метода, хотя Fry применяет его к другим объектам и в других предположениях.

В дальнейшем речь идет о вызовах, разговорах, абонентах и т. п. Эта терминология, взятая из телефонной практики, выбрана лишь для определенности и ни в какой мере не должна ограничивать собою области приложимости полученных результатов.

§ 1. Определения, обозначения и постановка задачи

Телефонистка, обслуживающая данную (многочисленную) группу абонентов, получает, в среднем, определенное число вызовов (требований) в час. Мы допускаем, что вероятность поступления данного числа k вызовов в данный промежуток времени t зависит только от длины этого промежутка, но не зависит ни от момента его начала, ни от того, сколько абонентов находилось в этот момент в состоянии ожидания разговора. Это условие и является предпосылкой стационарности очереди. Если вызов застает телефонистку занятой, абонент должен выждать, пока она переговорит со всеми абонентами, ранее него занявшими очередь. Длительность разговора между абонентом и телефонисткой предполагается случайной величиною; в дальнейшем через $f(t) dt$ обозначается вероятность того, что эта длительность заключена между t и $t + dt$ (таким образом, $f(t) dt$ есть относительное число разговоров, имеющих данную длительность). Средняя

*) Math. Zeitschr., т. 32, 1930, с. 64, стр. 729.

**) См. его книгу «Теория вероятностей для инженеров», ГТТИ, 1934.

продолжительность разговора равна

$$\mu_1 = \int_0^{\infty} t f(t) dt.$$

Если в единицу времени (за которую в телефонной практике обычно принимается час) происходит n вызовов, то $\mu_1 = \alpha$ есть среднее время занятости (нагрузка) телефонистки. Во всем дальнейшем предполагается $\alpha < 1$, так как в противном случае очередь, как легко видеть, безгранично растет и не может быть стационарной (такого рода установка была бы в качестве постоянной установки практически бессмысленной).

Таким образом данными нашей задачи служат: с одной стороны, функция $f(t)$ (закон распределения длительности разговоров), с другой, — число α (нагрузка телефонистки) или n (число вызовов в час). Ищется, в конечном счете, закон распределения времени ожидания, т. е. вероятность того, что абонент, производящий вызов в произвольно выбранный момент времени, должен будет прождать более чем t часов, прежде чем вступит в разговор с телефонисткой; наиболее важными показателями являются, однако, *среднее значение* этого времени ожидания и его *среднее квадратическое отклонение*.

В дальнейшем я обозначаю через π_k ($k=1, 2, \dots$) вероятность того, что при начале наудачу выбранного разговора *) число ожидающих очереди абонентов упадет с k до $k-1$; иначе говоря, π_k есть относительное число разговоров, начинающихся в указанных условиях. Через π_0 я подобным же образом буду обозначать относительное число разговоров, начинающихся без предварительного ожидания.

Очевидно, мы можем также сказать, что π_k ($k=0, 1, 2, \dots$) есть относительное число таких *окончаний* разговоров, в момент которых число занятых абонентов падает с $k+1$ до k ; в случае $k > 0$ это непосредственно ясно, так как

*) Идея рассмотрения моментов начала разговоров оказалась весьма плодотворной. О возникшей в связи с этим теории вложенных цепей Маркова см. работу Д. Кендалла «Стохастические процессы, встречающиеся в теории очередей, и их анализ методом вложенных цепей Маркова», сб. переводов «Математика», т. 3, вып. 6, 1959, стр. 97—111.— Б. Г.

момент окончания одного разговора есть, вместе с тем, момент начала следующего разговора; в случае же $k=0$ это следует из того, что за каждым окончанием в отсутствии ожидающих следует в точности одно начало без ожидания.

Но непосредственно очевидно, что число моментов, когда очередь переходит от $k+1$ участников к k , равно числу моментов обратного перехода от k к $k+1$; так как момент первого рода наступает только при окончании разговора, а момент второго рода — только при появлении нового вызова, то π_k ($k=0, 1, 2, \dots$) есть относительное число вызовов, в момент которых число занятых абонентов повышается с k до $k+1$.

Эти числа π_k мы будем называть *начальными индексами* нашей задачи; их роль будет чисто вспомогательной; однако, мы должны начать с их определения.

§ 2. Рекуррентное определение начальных индексов

В тех условиях, которые нами приняты относительно поступления вызовов, вероятность того, что в течение некоторого данного промежутка времени t произойдет k вызовов, выражается, как известно, формулой Пуассона*)

$$v_k(t) = e^{-\frac{\alpha}{\mu_1} t} \frac{\left(\frac{\alpha}{\mu_1} t\right)^k}{k!}.$$

Найдем прежде всего вероятность η_k того, что в течение данного, наудачу выбранного разговора произойдет k вызовов (иначе говоря, η_k есть относительное число тех разговоров, в течение которых происходит k вызовов); вероятность того, что выбранный разговор будет иметь длительность, заключенную между t и $t+dt$, равна по предположению $f(t)dt$; следовательно, по формуле полной вероятности:

$$\eta_k = \int_0^{\infty} v_k(t) f(t) dt.$$

*) Это есть общая теорема теории вероятностей, находящая себе применение не только в эксплуатации установок общего пользования, но, например, также в теории радиоактивного излучения. Доказательство можно найти, например, в книге Фрая «Теория вероятностей для инженеров».

Мы обозначим через π_0 вероятность того, что наудачу выбранный разговор начнется без предварительного ожидания; для того чтобы это случилось, необходимо и достаточно, чтобы предыдущий разговор начался в момент, когда никто не ждет (вероятность чего есть $\pi_0 + \pi_1$), и чтобы, сверх того, в течение этого разговора не последовало ни одного вызова (вероятность чего есть η_0). Отсюда

$$\pi_0 = (\pi_1 + \pi_0) \eta_0. \quad (1)$$

В случае $k > 0$ π_k есть вероятность того, что наудачу выбранный разговор начнется в момент, когда число ожидающих падает с k до $k - 1$; для того чтобы это случилось, необходимо и достаточно осуществление какой-нибудь одной из следующего ряда предпосылок:

1. Предыдущий разговор начался в момент, когда ожидающих не было (вероятность $\pi_0 + \pi_1$), и в течение этого разговора произошло k вызовов (вероятность η_k).

2. Предыдущий разговор начался в момент, когда число ожидающих уменьшилось с двух до одного (вероятность π_2), и в течение этого разговора произошло $k - 1$ вызовов (вероятность η_{k-1}).

.....

$k + 1$. Предыдущий разговор начался в момент, когда число ожидающих упало с $k + 1$ до k (вероятность π_{k+1}), и в течение этого разговора не последовало ни одного вызова (вероятность η_0).

По формуле полной вероятности мы, таким образом, получаем:

$$\pi_k = \pi_{k+1} \eta_0 + \pi_k \eta_1 + \pi_{k-1} \eta_2 + \dots + \pi_2 \eta_{k-1} + (\pi_1 + \pi_0) \eta_k \quad (k > 0). \quad (2)$$

Формулы (1) и (2) позволяют последовательно выразить все π_k через π_0 , а так как $\sum_{k=0}^{\infty} \pi_k = 1$, то и π_0 является определенным. Явные выражения этих начальных индексов через данные элементы нам в дальнейшем не понадобятся, но полученные рекуррентные формулы призваны играть основную роль.

§ 3. Основной закон стационарной очереди

Пусть $P(k)$ означает вероятность того, что в произвольно выбранный момент времени в точности k абонентов окажутся занятыми (т. е. что в случае $k > 0$ один говорит и $k - 1$ ожидают). Практически эта вероятность измеряется относительной длительностью того времени, в течение которого занято k абонентов.

Основным законом стационарной очереди назовем соотношение

$$P(k) = \pi_k \quad (k = 0, 1, 2, \dots)$$

справедливость которого и непосредственно довольно ясна; в самом деле, в § 2 мы видели, что π_k есть относительное число вызовов, встречающих при своем возникновении k занятых абонентов; очевидно, что это число при наших предположениях совпадает с относительной длительностью $P(k)$ того времени, в течение которого выполняется это условие.

Однако этот непосредственный вывод основного закона все же не является вполне убедительным; поэтому мы дадим формальное доказательство этого закона, тем более, что нужный для этого математический аппарат нам все равно понадобится в дальнейшем.

Итак, мы постараемся выразить величину $P(k)$ через начальные индексы π_i . С этой целью заметим следующее: вероятность того, что в произвольно выбранный момент времени телефонистка занята, равна α ; если это случилось, то вероятность того, что ведущийся разговор имеет длительность, заключенную между t и $t + dt$, равна относительной общей продолжительности таких разговоров, т. е.

$$\frac{t f(t) dt}{\mu_1};$$

если, наконец, разговор действительно имеет эту длительность, то вероятность $P_1(x < u < x + dx)$ того, что уже протекавшая до нашего вызова часть u этого разговора заключена между x и $x + dx$, составляет $\frac{dx}{t}$ (если $x < t$) и 0

(если $x > t$). Отсюда:

$$\begin{aligned} P(x < u < x + dx) &= \alpha \int_x^{\infty} \frac{t f(t) dt}{\mu_1} \frac{dx}{t} = \\ &= \frac{\alpha dx}{\mu_1} \int_x^{\infty} f(t) dt = \frac{\alpha}{\mu_1} [1 - F(x)] dx. \end{aligned}$$

Здесь левая часть есть вероятность того, что в произвольно выбранный момент времени мы застанем разговор, уже протекающий часть которого u заключена между x и $x + dx$, а в правой части через $F(x)$ обозначен интегральный закон распределения длительности разговоров. Вероятность же того, что в течение этой, уже протекающей, части разговора состоялось j вызовов, равна, очевидно

$$\omega_j = \int_0^{\infty} P(x < u < x + dx) v_j(x) = \frac{\alpha}{\mu_1} \int_0^{\infty} v_j(x) [1 - F(x)] dx.$$

Но, с другой стороны, очевидно, что

$$P(k) = \pi_k \omega_0 + \pi_{k-1} \omega_1 + \dots + (\pi_1 + \pi_0) \omega_{k-1},$$

откуда

$$P(k) = \frac{\alpha}{\mu_1} \int_0^{\infty} [1 - F(x)] u_{k-1}(x) dx, \quad (3)$$

где положено:

$$u_{k-1}(x) = \pi_k v_0(x) + \pi_{k-1} v_1(x) + \dots + (\pi_1 + \pi_0) v_{k-1}(x).$$

Функции $u_k(x)$, как легко непосредственно проверить, удовлетворяют разностно-дифференциальному уравнению

$$u'_k(x) = \frac{\alpha}{\mu_1} [u_{k-1}(x) - u_k(x)];$$

поэтому, написав соотношение (2) в виде:

$$\pi_k = \int_0^{\infty} u_k(x) f(x) dx,$$

и интегрируя по частям, мы находим:

$$\begin{aligned}\pi_k &= u_k(0) + \int_0^{\infty} u'_k(x) [1 - F(x)] dx = \\ &= \pi_{k+1} + \frac{\alpha}{\mu_1} \int_0^{\infty} [u_{k+1}(x) - u_k(x)] [1 - F(x)] dx = \\ &= \pi_{k+1} + P(k) - P(k+1),\end{aligned}$$

откуда

$$\pi_k - P(k) = \pi_{k+1} - P(k+1) = \text{const.}$$

Но π_k и $P(k)$ — члены сходящихся рядов, в силу чего эта постоянная должна обращаться в нуль, и таким образом, основной закон доказан.

В силу этого закона рекуррентные формулы (1) и (2) имеют место для величин $P(k)$ и позволяют в случае необходимости их последовательно определять; однако, получающееся отсюда общее выражение для $P(k)$ неудобно для расчетов и для теоретического употребления, вследствие чего мы предпочитаем в дальнейшем пользоваться соотношениями (1) и (2) в их первоначальном виде:

$$P(k) = \int_0^{\infty} u_k(x) f(x) dx. \quad (4)$$

§ 4. Определение среднего времени ожидания

Время ожидания γ (случайная величина), т. е. время, протекающее от момента вызова до начала разговора, складывается из двух составляющих: 1) время τ от момента вызова до окончания того разговора, который ведется в момент вызова, и 2) суммарная длительность T разговоров всех тех абонентов, которые в момент произведенного вызова находились в состоянии ожидания (разумеется, $T=0$, если вызов произошел в момент, когда нет ожидающих, и $T=\tau=\gamma=0$, если вызов застал телефонистку свободной).

Таким образом:

$$\gamma = \tau + T,$$

где τ и T — также случайные величины. Для метода последующих расчетов важно иметь в виду следующее замечание.

Очень легко найти закон распределения для величины τ ; зная величины $P(k)$, найденные нами в § 3, столь же нетрудно найти закон распределения величины T ; однако, это не дало бы нам возможности написать закон распределения для величины γ , потому что величины τ и T *взаимно зависимы*, и в этом — главная трудность проблемы. В самом деле, если, например, τ мало, то это есть при прочих равных условиях указание на то, что разговор длится уже относительно долго; это же, в свою очередь, указывает на то, что, вероятно, скопилось много ожидающих, так что T получает шансы стать большим. Каким образом может быть преодолено это затруднение, мы увидим в дальнейшем; здесь же, где речь идет только о математическом ожидании величины γ , это обстоятельство не играет, конечно, никакой роли, так как (обозначая математическое ожидание случайной величины той же буквой, но с чертой сверху):

$$\bar{\gamma} = \bar{\tau} + \bar{T}.$$

Чтобы найти $\bar{\tau}$, заметим, что вероятность застать разговор длительности, заключенной между t и $t + dt$, равна, как мы уже видели,

$$\frac{\alpha}{\mu_1} t f(t) dt;$$

при этом же условии математическое ожидание величины τ , очевидно, равно $\frac{1}{2} t$; отсюда

$$\bar{\tau} = \frac{\alpha}{2\mu_1} \int_0^{\infty} t^2 f(t) dt = \frac{\alpha\mu_2}{2\mu_1},$$

где μ_2 означает математическое ожидание квадрата длительности разговора.

Что касается величины T , то при условии, что в момент вызова занято $j \geq 1$ абонентов, ее математическое ожидание, очевидно, становится равным $(j-1)\mu_1$; отсюда:

$$\bar{T} = \mu_1 \sum_{j=1}^{\infty} (j-1) P(j);$$

а так как $\sum_{j=1}^{\infty} j P(j)$ есть математическое ожидание числа занятых абонентов в произвольно выбранный момент времени,

очевидно, совпадающее с суммарным временем занятости всех абонентов в течение часа, то мы имеем:

$$\sum_{j=1}^{\infty} j P(j) = n(\mu_1 + \bar{\gamma}) = \alpha + \frac{\alpha}{\mu_1} \bar{\gamma};$$

с другой стороны, очевидно

$$\sum_{j=1}^{\infty} P(j) = 1 - P(0) = \alpha,$$

так что

$$\bar{T} = \mu_1 \left(\alpha + \frac{\alpha}{\mu_1} \bar{\gamma} - \alpha \right) = \alpha \bar{\gamma};$$

таким образом

$$\bar{\gamma} = \frac{\alpha}{2} \frac{\mu_2}{\mu_1} + \alpha \bar{\gamma},$$

откуда

$$\bar{\gamma} = \frac{\alpha}{2(1-\alpha)} \frac{\mu_2}{\mu_1}. \quad (5)$$

Эта простая формула полностью решает вопрос о среднем времени ожидания *); мы видим, что $\bar{\gamma}$ зависит не только от средней длительности разговора, но и от дисперсии. Практически это лишний раз подчеркивает важность стандартизации разговора между абонентом и телефонисткой.

В качестве числового примера рассмотрим случай нулевой дисперсии ($\mu_2 = \mu_1^2$), в налаженной телефонной практике близкий к действительности; положим $\alpha = 0,6$, $\mu_1 = 8$ сек.; это дает $\bar{\gamma} = 6$ сек.; если при тех же α и μ_1 предположить распределение длительности разговоров по показательному закону (ввиду его простоты часто, хотя и без достаточных оснований принимаемому за базу теоретических изысканий), то $\mu_2 = 2\mu_1^2$, и среднее время ожидания увеличивается вдвое: $\bar{\gamma} = 12$ сек.

*) Мы считаем нужным указать, что имеющая некоторое распространение среди практиков формула д-ра Маттиаса неверна, и вывод ее основан на ошибочных предположениях.

§ 5. Формулы для основных вероятностей

Мы видели, что среднее время ожидания $\bar{\gamma}$ находится вполне элементарно; однако, уже математическое ожидание величины γ^2 этим методом не удалось бы установить ввиду отмеченной зависимости между величинами τ и T . Здесь необходим более глубокий анализ.

Прежде всего, поставим следующую задачу: если известно, что в некоторый момент времени происходит разговор, длительность которого равна t , то какова вероятность того, что в этот момент занято j абонентов, и что в то же время уже протекающая часть разговора u заключена между x и $x + dx$? Обозначая эту вероятность через $P_t(j, x < u < x + dx)$, мы, очевидно, будем иметь в легко понятных обозначениях:

$$P_t(j, x < u < x + dx) = P_t(j) P_{t,j}(x < u < x + dx) = \\ = P_t(x < u < x + dx) = P_{t, x < u < x + dx}(j).$$

Правая часть этих равенств легко может быть вычислена; в самом деле, во-первых (ср. § 3):

$$P_t(x < u < x + dx) = \begin{cases} \frac{dx}{t}, & \text{если } x < t, \\ 0, & \text{если } x > t; \end{cases}$$

во-вторых, последний множитель совпадает с $P_{x < u < x + dx}(j)$ (так как при фиксированной протекающей части разговора его общая длительность, очевидно, не может влиять на степень скученности в данный момент), и следовательно, равен:

$$\pi_j v_0(x) + \pi_{j-1} v_1(x) + \dots + (\pi_1 + \pi_0) v_{j-1}(x) = u_{j-1}(x).$$

Таким образом:

$$P_t(j) P_{t,j}(x < u < x + dx) = u_{j-1}(x) \frac{dx}{t}$$

в случае $x < t$ и 0 в случае $x > t$.

Но при данном t неравенства $x < u < x + dx$ эквивалентны неравенствам $t - x - dx < \tau < t - x$, а потому:

$$P_t(j) P_{t,j}(x < \tau < x + dx) = u_{j-1}(t - x) \frac{dx}{t}$$

при $x < t$ и 0 при $x > t$. Чтобы получить отсюда величину $P(j) P_j(x < \tau < x + dx)$, т. е. вероятность того, что

в произвольно выбранный момент времени мы найдем j занятых абонентов и что застигнутый нами при этом разговор продлится еще в течение времени, заключенного между x и $x + dx$, надо, очевидно, помножить найденную вероятность на $\frac{\alpha}{\mu_1} t f(t) dt$ и проинтегрировать по t от x до ∞ (ибо при $t < x$ подынтегральная функция равна нулю). Таким образом:

$$P(j) P_j(x < \tau < x + dx) = \frac{\alpha}{\mu_1} dx \int_x^{\infty} u_{j-1}(t-x) f(t) dt, \quad (6)$$

где $j = 1, 2, 3, \dots$

§ 6. Характеристическая функция времени ожидания

Как известно, для изучения всех вопросов, связанных с законом распределения случайной величины γ , в частности, для вычисления ее моментов любого порядка, удобнее всего найти ее характеристическую функцию, т. е. математическое ожидание выражения $e^{i\tau\xi}$, рассматриваемое как функция параметра ξ . Так мы и поступим. Пользуясь для обозначения математического ожидания символом E , мы будем иметь:

$$\varphi(\xi) = E(e^{i\tau\xi}) = E(e^{i\tau\xi} e^{i\tau\xi}).$$

Однако математическое ожидание произведения в данном случае не равно произведению математических ожиданий, ввиду отмеченной зависимости величин τ и T . Поэтому для отыскания характеристической функции величины γ мы должны прибегнуть к более сложному приему.

Если мы условимся обозначать через E_j математические ожидания, вычисленные в предположении, что вызов застал j занятых абонентов, то в силу известных правил теории вероятностей:

$$\varphi(\xi) = E(e^{i\tau\xi}) = \sum_{j=0}^{\infty} P(j) E_j(e^{i\tau\xi}) = \sum_{j=0}^{\infty} P(j) E_j(e^{i\tau\xi} e^{i\tau\xi}).$$

Но при данном значении j величины τ и T взаимно независимы, так как, очевидно, длительность разговоров $j-1$ ожидающих абонентов не зависит от того, сколько еще времени будет длиться застигнутый разговор (величина T

только потому зависела от τ , что j зависело от τ). Поэтому:

$$\begin{aligned}\Phi(\xi) &= \sum_{j=0}^{\infty} P(j) E_j(e^{i\tau\xi}) E_j(e^{iT\xi}) = \\ &= P(0) + \sum_{j=1}^{\infty} P(j) E_j(e^{i\tau\xi}) E_j(e^{iT\xi}).\end{aligned}$$

После достигнутого таким образом разделения величин τ и T дальнейший метод вычисления уже ясен. Прежде всего, легко вычисляется $E_j(e^{iT\xi})$. В самом деле, если мы обозначим через $\psi(\xi)$ характеристическую функцию длительности разговора, т. е. положим

$$\psi(\xi) = \int_0^{\infty} e^{iT\xi} f(t) dt,$$

и заметим, что T есть суммарная длительность $j-1$ взаимно независимых разговоров, то мы без всяких вычислений можем написать:

$$E_j(e^{iT\xi}) = [\psi(\xi)]^{j-1} \quad (j=1, 2, \dots).$$

Далее, в силу формулы (6):

$$\begin{aligned}P(j) E_j(e^{i\tau\xi}) &= \int_0^{\infty} P(j) e^{ix\xi} P_j(x < \tau < x+dx) = \\ &= \frac{\alpha}{\mu_1} \int_0^{\infty} e^{ix\xi} dx \int_x^{\infty} u_{j-1}(t-x) f(t) dt \\ &\quad (j=1, 2, \dots),\end{aligned}$$

откуда

$$\Phi(\xi) = P(0) + \frac{\alpha}{\mu_1} \int_0^{\infty} e^{ix\xi} dx \int_x^{\infty} f(t) \left\{ \sum_{j=0}^{\infty} [\psi(\xi)]^j u_j(t-x) \right\} dt. \quad (7)$$

Пользуясь выражением функции $u_j(x)$, очень легко показать, что

$$\sum_{j=0}^{\infty} \alpha^j u_j(z) = e^{\frac{z}{\mu_1} (\alpha-1)} \left\{ P(0) + \sum_{k=1}^{\infty} \alpha^{k-1} P(k) \right\}, \quad (8)$$

откуда в силу формулы (7)

$$\varphi(\xi) = P(0) + \frac{\alpha}{\mu_1} \chi(\xi) \int_x^{\infty} e^{ix\xi} dx \int_x^{\infty} e^{\frac{\alpha}{\mu_1} [\psi(\xi) - 1] (t-x)} f(t) dt,$$

где положено:

$$\chi(\xi) = P(0) + \sum_{k=1}^{\infty} [\psi(\xi)]^{k-1} P(k).$$

Полагая

$$\frac{\alpha}{i\mu_1} [\psi(\xi) - 1] = \eta.$$

находим:

$$\begin{aligned} \varphi(\xi) &= P(0) + \frac{\alpha}{\mu_1} \chi(\xi) \int_0^{\infty} e^{ix(\xi-1)} dx \int_x^{\infty} e^{it\eta} f(t) dt = \\ &= P(0) + \frac{\alpha}{\mu_1} \chi(\xi) \int_0^{\infty} e^{it\eta} f(t) dt \int_0^t e^{ix(\xi-\eta)} dx = \\ &= P(0) + \frac{\alpha}{\mu_1} \frac{\chi(\xi)}{(\eta-\xi)i} \int_0^{\infty} e^{it\eta} f(t) [1 - e^{it(\xi-\eta)}] dt = \\ &= P(0) + \frac{\alpha}{i\mu_1} \chi(\xi) \frac{\psi(\eta) - \psi(\xi)}{\eta - \xi} \end{aligned}$$

Но в силу формул (4) и (8):

$$\begin{aligned} \chi(\xi) &= P(0) + \frac{1}{\psi(\xi)} \int_0^{\infty} f(x) \left\{ \sum_{k=0}^{\infty} [\psi(\xi)]^k u_k(x) - P(0) \right\} dx = \\ &= P(0) \frac{\psi(\xi) - 1}{\psi(\xi)} + \frac{\chi(\xi)}{\psi(\xi)} \int_0^{\infty} e^{\frac{\alpha}{\mu_1} [\psi(\xi) - 1] x} f(x) dx = \\ &= P(0) \frac{\psi(\xi) - 1}{\psi(\xi)} + \chi(\xi) \frac{\psi(\eta)}{\psi(\xi)}, \end{aligned}$$

откуда

$$\chi(\xi) = P(0) \frac{\psi(\xi) - 1}{\psi(\xi) - \psi(\eta)},$$

и следовательно:

$$\begin{aligned}\varphi(\xi) &= P(0) \left\{ 1 - \frac{\alpha}{i\mu_1} \frac{\psi(\xi) - 1}{\eta - \xi} \right\} = \\ &= (1 - \alpha) \left\{ 1 - \frac{\alpha [\psi(\xi) - 1]}{\alpha [\psi(\xi) - 1 - i\mu_1 \xi]} \right\},\end{aligned}$$

или

$$\varphi(\xi) = \frac{1 - \alpha}{1 - \alpha \left\{ \frac{\psi(\xi) - 1}{i\mu_1 \xi} \right\}}.$$

Эта замечательная по своей простоте формула позволяет с большою легкостью находить выражения моментов величины γ через моменты длительности разговора; для этого надо только вычислять последовательные производные $\varphi(\xi)$ и, деля их на соответствующие степени i , полагать в них $\xi = 0$. Так, $\frac{1}{i} \varphi'(0) = \bar{\gamma}$ дает нам прежнюю формулу (5).

Далее,

$$-\varphi''(0) = E(\gamma^2) = 2\bar{\gamma}^2 + \frac{\alpha\mu_2}{3(1-\alpha)\mu_1},$$

где

$$\mu_2 = \int_0^{\infty} t^2 f(t) dt.$$

Отсюда дисперсия:

$$\sigma^2(\gamma) = E(\gamma^2) - \bar{\gamma}^2 = \bar{\gamma}^2 + \frac{\alpha\mu_2}{3(1-\alpha)\mu_1}. \quad (9)$$

Эта формула показывает, что среднее квадратическое отклонение величины γ всегда превышает ее среднее значение; этим хорошо объясняется то явление, что даже при небольшом среднем времени ожидания мы часто имеем дело со сравнительно большими задержками. Так, например, при условии стандартной длительности разговора 8 сек. и при $\alpha = 0,6$ мы имели $\bar{\gamma} = 6$ сек. При тех же условиях формула (9) дает $\sigma(\gamma) > 8$ сек., так что при среднем времени ожидания 6 сек. не редкостью будет задержка в 15 сек. и более, что является уже ощутительным.

О СРЕДНЕМ ВРЕМЕНИ ПРОСТОЯ СТАНКОВ*)

1. Решаемая в настоящей статье общая задача имеет значение для всех тех отраслей промышленности, где наблюдению одного рабочего одновременно поручается несколько станков (или других установок) одинакового типа, причем функция рабочего состоит в выполнении операций, необходимость в которых для каждого отдельного станка возникает в случайном порядке (ликвидация неполадок, мелкий ремонт и т. п.).

Момент, в который станок начинает требовать той или иной операции, мы будем называть моментом его остановки. Время, требующееся для совершения той или другой операции, может быть различным для различных операций и допускать какие угодно вариации для одной и той же операции.

В дальнейшем мы принимаем следующие обозначения:

λ — число станков, поручаемых одному рабочему;

$x dt$ — вероятность того, что станок, работающий в момент t , остановится ранее момента $t + dt$ (dt предполагается весьма малым);

$f(t) dt$ — относительное число операций, для которых потребное время заключено между t и $t + dt$.

Эти три элемента являются данными. Ищется среднее время ожидания γ , т. е. среднее значение промежутка времени, протекающего от момента остановки до начала соответствующей операции; при этом предполагается, что рабочий обслуживает станки в порядке очереди, т. е. в порядке их остановок.

В качестве вспомогательных величин нам понадобятся следующие:

*) Матем. сб., т. 40 (1933), № 2, стр. 119—123.

μ — средняя длительность одной операции, n — число остановок одного станка в единицу времени;

$\varphi(t)$ — вероятность того, что станок, работающий в момент t_0 , не остановится до момента $t_0 + t$;

π_k — относительное число остановок, происходящих в такие моменты, когда число уже стоящих станков равно k .

Все эти величины могут быть определены из данных нашей задачи.

Следует заметить, что формулами, которые даются в настоящей статье, целесообразно пользоваться лишь для небольших значений λ (в пределах 1—2 десятков). При большем числе станков можно с успехом применять, в порядке приближения, значительно более простые методы расчета, которые нами установлены в другой статье*).

2. Единицу времени, которая может быть выбрана произвольно, целесообразно принять весьма большою; для краткости условимся называть ее часом.

Для средней длительности операции мы, очевидно, имеем формулу:

$$\mu = \int_0^{\infty} t f(t) dt.$$

Определим, далее, функцию $\varphi(t)$; для этого заметим, что, очевидно,

$$\varphi(t + dt) = \varphi(t) (1 - \kappa dt),$$

откуда

$$\varphi'(t) = -\kappa \varphi(t),$$

или

$$\varphi(t) = C e^{-\kappa t},$$

причем очевидное условие $\varphi(0) = 1$ дает $C = 1$, так что

$$\varphi(t) = e^{-\kappa t}.$$

Вероятность же того, что станок, работающий в начале некоторого промежутка длины t , остановится в течение этого промежутка, равна

$$1 - \varphi(t) = 1 - e^{-\kappa t},$$

*) «Математическая теория стационарной очереди», Матем. сб., т. 39 (1932), № 4, стр. 73.

и далее, вероятность того, что для станка, работающего в момент t_0 , первая остановка последует в течение малого промежутка времени $(t_0 + t, t_0 + t + dt)$, равна

$$\varphi(t) \kappa dt = \kappa e^{-\kappa t} dt;$$

отсюда мы можем определить среднее значение рабочего периода станка, т. е. времени, протекающего от момента пуска (конца операции) до ближайшей остановки:

$$\int_0^{\infty} t \kappa e^{-\kappa t} dt = \frac{1}{\kappa}.$$

Заметим, что практически величину κ лучше всего находить именно из этой формулы, так как средний рабочий период станка легко поддается непосредственному измерению.

3. Теперь мы можем перейти к установлению формулы, определяющей среднее время ожидания γ . С этой целью заметим, что рабочий час для каждого станка складывается из чередующихся периодов трех различных категорий: рабочий период (средней длины $\frac{1}{\kappa}$), период ожидания (средней длины γ) и период операции (средней длины μ). Так как мы обозначили через n число остановок станка в течение часа, то

$$n \left(\frac{1}{\kappa} + \gamma + \mu \right) = 1. \quad (1)$$

В этом уравнении, кроме γ , неизвестно еще и n . Поэтому мы должны искать второе соотношение, связывающее те же неизвестные. С этой целью заметим, что для рабочего его рабочий час складывается из чередующихся периодов двух категорий — периодов занятости и периодов свободы. Сумму длин всех периодов занятости легко найти. Число станков равно λ , каждый станок требует в час n операций, средняя длительность операции равна μ . Отсюда время загрузки рабочего равно $n\lambda\mu$. Сложнее найти время, в течение которого рабочий свободен. Заметим прежде всего, что число моментов освобождения рабочего, очевидно, равно числу моментов, когда он от свободного состояния переходит к работе (так как моменты этих двух категорий строго чередуются). Но второе число есть, вместе с тем, число

остановок, происходящих в периоды, когда все остальные станки работают (и потому рабочий свободен). Согласно нашим обозначениям это число есть $\lambda n \pi_0$; таково, следовательно, число периодов свободы для рабочего. Что касается средней длительности такого периода, то она может быть найдена из следующих соображений. Если в данный момент все станки работают, то вероятность того, что и по истечении времени t все они еще будут работать, равна $e^{-\lambda x t}$; вероятность же того, что какой-нибудь из них остановится в промежуток времени $(t, t + dt)$, равна (пренебрегая малыми высших порядков) $\lambda x dt$. Следовательно, вероятность того, что длительность данного периода свободы для рабочего будет заключена между t и $t + dt$, равна $e^{-\lambda x t} \lambda x dt$, а значит, средняя длительность периода свободы рабочего есть

$$\int_0^{\infty} t e^{-\lambda x t} \lambda x dt = \frac{1}{\lambda x}.$$

Отсюда сумма длин всех периодов свободы равна

$$\lambda n \pi_0 \cdot \frac{1}{\lambda x} = n \frac{\pi_0}{x},$$

и мы получаем соотношение

$$n \left(\lambda \mu + \frac{\pi_0}{x} \right) = 1. \quad (2)$$

Исключая n из уравнений (1) и (2), мы находим:

$$\gamma = (\lambda - 1) \mu - \frac{1 - \pi_0}{x}.$$

Этот результат полностью решал бы поставленную задачу, если бы π_0 было нам известно. Определение величины π_0 , которому будет посвящено все дальнейшее, и составляет в сущности, математическое содержание предлагаемого решения, так как все предшествующее было элементарным расчетом.

4. Число π_k , представляющее собою, по определению, относительное число остановок, в момент которых число стоящих станков увеличивается с k до $k + 1$, очевидно, есть, вместе с тем, относительное число таких операций, в конце которых число стоящих станков уменьшается

с $k+1$ до k . Обозначим через $\eta_{k,j}$ вероятность того, что в течение операции, в начале которой число стоящих станков равно k , остановится еще j станков. Для того чтобы некоторая операция началась без предварительного ожидания (вероятность π_0), необходимо и достаточно, чтобы: 1) предшествующая операция началась в отсутствии ожидающих (вероятность $\pi_1 + \pi_0$) и 2) в течение этой операции ни один станок не остановился (вероятность $\eta_{1,0}$). Отсюда:

$$\pi_0 = (\pi_0 + \pi_1) \eta_{1,0}. \quad (3)$$

Подобным же образом, если $0 < k < \lambda - 1$, то для того, чтобы какая-либо операция началась в момент, когда число стоящих станков уменьшается с $k+1$ до k^* (вероятность π_k), необходимо и достаточно, чтобы: либо предшествующая операция началась в момент, когда число стоящих станков уменьшается с $k+2$ до $k+1$ (вероятность π_{k+1}), и в течение этой операции не произошло ни одной остановки (вероятность $\eta_{k+1,0}$); либо предшествующая операция началась в момент, когда число стоящих станков уменьшается с $k+1$ до k (вероятность π_k), и в течение этой операции остановился один станок (вероятность $\eta_{k,1}$), и т. д.; либо, наконец, чтобы предшествующая операция началась в отсутствии ожидающих (вероятность $\pi_1 + \pi_0$) и в течение этой операции произошло k остановок (вероятность $\eta_{1,k}$). Таким образом мы находим:

$$\pi_k = \pi_{k+1} \eta_{k+1,0} + \pi_k \eta_{k,1} + \dots + (\pi_1 + \pi_0) \eta_{1,k} \quad (0 < k < \lambda - 1). \quad (4)$$

Наконец, в случае $k = \lambda - 1$ мы аналогичным рассуждением находим:

$$\pi_{\lambda-1} = \pi_{\lambda-1} \eta_{\lambda-1,1} + \dots + (\pi_1 + \pi_0) \eta_{1,\lambda-1}. \quad (5)$$

Если предполагать все величины $\eta_{k,j}$ известными, то, очевидно, из уравнений (3) и (4) можно последовательно выразить все π_k линейно через π_0 ; уравнение (5) представляет собою следствие соотношений (3) и (4) и новых результатов

*) Заметим, что при $k > 0$ момент начала данной операции есть, вместе с тем, момент окончания предшествующей операции.

не дает, но очевидное соотношение

$$\sum_{k=0}^{\lambda-1} \pi_k = 1$$

позволяет определить и π_0 .

Таким образом, задача сводится к определению коэффициентов $\eta_{k,j}$, что может быть сделано без затруднений. В самом деле, если данная операция длится t часов и если в начале этой операции стоят k станков, то вероятность того, что в течение этой операции из $\lambda - k$ работающих станков остановятся ровно j , очевидно, есть

$$C'_{\lambda-k} (1 - e^{-xt})^j (e^{-xt})^{\lambda-k-j}.$$

А так как вероятность того, что данная операция будет иметь длительность, заключенную между t и $t + dt$, есть $f(t) dt$, то

$$\eta_{k,j} = C'_{\lambda-k} \int_0^{\infty} (1 - e^{-xt})^j (e^{-xt})^{\lambda-k-j} f(t) dt,$$

что и решает задачу.

5. Зная среднее время простоя, мы без труда можем решать все вопросы, связанные с наиболее выгодным распределением станков между рабочими. Для практических расчетов, основанных на изложенной теории, необходимо, разумеется, создать специальный аппарат в виде номограммы или таблицы, что, вероятно, не представит затруднений. Теоретически же задача приведенными нами соображениями решается до конца.

ПОТОКИ СЛУЧАЙНЫХ СОБЫТИЙ БЕЗ ПОСЛЕДЕЙСТВИЯ*)

§ 1. Постановка задачи

Поток однородных случайных событий представляет собой случайный процесс $x(t)$ ($t \geq 0$), в котором функция $x(t)$ (число событий, наступающих в промежутке времени**) $(0, t)$ — неубывающая и способна принимать лишь целые неотрицательные значения. Полное стохастическое описание такого потока мы получаем, задавая для любого натурального числа n и для любой группы из n моментов времени $0 < t_1 < t_2 < \dots < t_n$ n -мерный закон распределения вектора $[x(t_1), x(t_2), \dots, x(t_n)]$. Вместо этого можно, очевидно, задать и закон распределения вектора $[x(t_1), x(t_2) - x(t_1), \dots, x(t_n) - x(t_{n-1})]$.

Поток событий называется *стационарным*, если закон распределения вектора $[x(a + t_i) - x(a), 1 \leq i \leq n]$ зависит только от чисел $t_1, t_2, \dots, t_n$, но не зависит от a (и, следовательно, при любом $a \geq 0$ совпадает с законом распределения вышеприведенного вектора $x(t_i), 1 \leq i \leq n$). Поток называется *поток без последействия*, если для любой системы (u_i, v_i) ($1 \leq i \leq n$) попарно не пересекающихся промежутков времени разности $x(v_i) - x(u_i)$ ($1 \leq i \leq n$) представляют собой взаимно независимые случайные величины. Если $v_k(\alpha, \beta)$ ($0 \leq \alpha < \beta, k \geq 0$ целое) означает вероятность равенства $x(\beta) - x(\alpha) = k$ (т. е. вероятность того, что в течение промежутка (α, β) наступит k событий данного потока), то поток без последействия, очевидно, полностью описывается заданием системы функций $v_k(\alpha, \beta)$ ($0 \leq \alpha < \beta$,

*) Теория вероятностей и ее применения, т. 1, вып. 1, 1956, стр. 3—17.

**) Промежутком времени (a, b) мы во всем дальнейшем называем совокупность всех t , для которых $a \leq t < b$.

$k \geq 0$), так как тем самым задается распределение разности $x(\beta) - x(\alpha)$ для любого промежутка (α, β) . Если данный поток к тому же стационарный, то величина $v_k(\alpha, \alpha + t)$ зависит только от k и t и может быть обозначена через $v_k(t)$ (вероятность наступления k событий в произвольном промежутке длины t).

Если к требованиям стационарности и отсутствия последействия добавить еще условие

$$\lim_{t \rightarrow 0} \frac{1 - v_0(t) - v_1(t)}{t} = 0, \quad (1)$$

то, как известно ([1], глава 1),

$$v_k(t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!} \quad (k = 0, 1, \dots);$$

распределение величины $x(t)$ при сделанных трех допущениях определяется, таким образом, однозначно с точностью до произвольного положительного числа λ (это число называют интенсивностью данного потока).

Недавно Редхеффером [2] и автором ([1], § 8) найдена общая форма стационарных потоков без последействия без дополнительного требования (1). С другой стороны, уже давно известна форма потока без последействия, удовлетворяющего требованию (1), но, вообще говоря, нестационарного*).

За последнее время, с одной стороны, обнаружилось, что для приложений имеет значение исследование потоков без последействия весьма общего вида. С другой стороны, удалось найти методы, сделавшие возможным исследования такого рода для некоторых специальных задач (см., например, рассмотрение известной задачи Эрланга у Пальма [3] или у автора [1], § 24; см. также работу Такача [4] о дробовом эффекте). Поэтому естественно встает вопрос об общей форме потоков без последействия. Настоящая работа имеет целью дать ответ на этот вопрос; полученное решение задачи можно признать удовлетворительным, так как все элементы произвола искомой структуры выявлены в нем с исчерпывающей ясностью.

*) Правда, при этом обычно предполагается еще существование некоторой «мгновенной интенсивности» $\lambda(t)$ потока любого момента времени t .

Пусть $\Lambda(t)$ означает (конечное или бесконечное) математическое ожидание числа событий данного потока, наступающих в промежутке $(0, t)$. Очевидно, $\Lambda(t)$ есть неотрицательная и неубывающая функция от t , которую мы будем называть *ведущей функцией* данного потока. Будем называть этот поток *финитным*, если $\Lambda(t) < +\infty$ ($t \geq 0$). Во всем дальнейшем мы будем рассматривать только финитные потоки. Случай $\Lambda(t) = +\infty$ представляет собой задачу совсем особого рода, вряд ли способную отвечать запросам большинства приложений.

Если $\Lambda(t)$ — любая неотрицательная, конечная, неубывающая функция от t , то формулы

$$v_k(\alpha, \beta) = \exp \left\{ - [\Lambda(\beta) - \Lambda(\alpha)] \right\} \frac{[\Lambda(\beta) - \Lambda(\alpha)]^k}{k!} \quad (2)$$

$$(0 \leq \alpha < \beta, k \geq 0)$$

всегда определяют некоторый поток без последствий, ведущая функция которого совпадает с $\Lambda(t)$. Именно этот поток обычно и рассматривают, когда исходят от ведущей функции $\Lambda(t)$. Однако такое ограничение, вообще говоря, представляется необоснованным. Как мы увидим, потоки типа (2) образуют лишь частный случай финитных потоков без последствий, так что, ограничиваясь потоками типа (2), мы подменили бы общее исследование проблемы рассмотрением частных случаев. Правда, потоки типа (2) образуют собою, как мы увидим, в известном смысле фундамент всей совокупности финитных потоков без последствий; но это как раз и надо показать с полной ясностью для того, чтобы специальное исследование потоков типа (2) получило хотя бы временное обоснование.

§ 2. Леммы о финитных потоках

Для решения поставленной задачи нам понадобится ряд элементарных и по большей части легко доказуемых вспомогательных предложений. Многие из этих лемм не связаны с отсутствием последствий и имеют силу для всех финитных потоков; такими леммами мы и займемся в настоящем параграфе; во всех последующих предложениях этого параграфа данный поток предполагается финитным, а во всем остальном — произвольным.

Лемма 1. При $0 \leq \alpha < \beta$ всегда $\sum_{k=0}^{\infty} v_k(\alpha, \beta) = 1$.

Доказательство. Очевидно, разность $1 - \sum_{i=0}^k v_i(\alpha, \beta)$ есть вероятность того, что в промежутке (α, β) наступит более k событий данного потока. Поэтому на основании неравенства Чебышева

$$1 - \sum_{i=0}^k v_i(\alpha, \beta) < \frac{\Lambda(\beta) - \Lambda(\alpha)}{k},$$

и, следовательно,

$$1 - \sum_{i=0}^k v_i(\alpha, \beta) \rightarrow 0 \quad (k \rightarrow \infty),$$

ч. т. д.

Лемма 2. При $k \geq 0$ и $t > 0$ существует предел

$$\lim_{h \rightarrow +0} v_k(t-h, t+h) = p_k(t). \quad (3)$$

Доказательство. Положим для $k \geq 0$, $0 \leq \alpha < \beta$

$$\sum_{r=k}^{\infty} v_r(\alpha, \beta) = \psi_k(\alpha, \beta).$$

Так как $\psi_k(\alpha, \beta)$ есть вероятность того, что в промежутке (α, β) наступит по меньшей мере k событий, то величина $\psi(t-h, t+h)$ при положительном и убывающем h не может возрастать. Поэтому всегда существует

$$\lim_{h \rightarrow +0} \psi_k(t-h, t+h).$$

В силу соотношения $v_k(\alpha, \beta) = \psi_k(\alpha, \beta) - \psi_{k+1}(\alpha, \beta)$ существует и предел (3), ч. т. д.

Условимся называть точку (момент времени) t *регулярной* или *сингулярной* точкой данного финитного потока, смотря по тому, будет ли $p_0(t) = 1$ или $p_0(t) < 1$.

Лемма 3. Все точки непрерывности функции $\Lambda(t)$ — регулярные, а все ее точки разрыва — сингулярные точки данного потока.

Доказательство. Так как

$$1 - v_0(t-h, t+h) = \sum_{k=1}^{\infty} v_k(t-h, t+h) \leq \\ \leq \sum_{k=1}^{\infty} k v_k(t-h, t+h) = \Lambda(t+h) - \Lambda(t-h),$$

то в случае $\Lambda(t+0) = \Lambda(t-0)$ мы имеем

$$1 - p_0(t) = \lim_{h \rightarrow +0} [1 - v_0(t-h, t+h)] = 0,$$

что доказывает первое утверждение леммы 3.

Далее, при $t > 0$, $0 < h < t$ и любом целом $m > 0$

$$\Lambda(t+h) - \Lambda(t-h) = \sum_{k=1}^{\infty} k v_k(t-h, t+h) = \\ = \sum_{k=1}^{\infty} \psi_k(t-h, t+h) = \sum_{k=1}^m \psi_k(t-h, t+h) + \\ + \sum_{k>m} \psi_k(t-h, t+h).$$

Пусть $\varepsilon > 0$ задано произвольно. Закрепим сначала некоторое h_0 , $0 < h_0 < t$, и выберем число m столь большим, чтобы было

$$R_m(h_0) = \sum_{k>m} \psi_k(t-h_0, t+h_0) < \varepsilon;$$

тогда и подавно при $0 < h \leq h_0$

$$R_m(h) < \varepsilon,$$

и, следовательно,

$$\Lambda(t+h) - \Lambda(t-h) < \sum_{k=1}^m \psi_k(t-h, t+h) + \varepsilon \\ (0 < h \leq h_0). \quad (4)$$

Но здесь

$$\sum_{k=1}^m \psi_k(t-h, t+h) \leq m \psi_1(t-h, t+h) \leq \\ \leq m [1 - v_0(t-h, t+h)].$$

Если $p_0(t) = 1$ (т. е. t — регулярная точка потока), то

$$1 - v_0(t-h, t+h) \rightarrow 0 \quad (h \rightarrow +0);$$

поэтому при достаточно малом $h > 0$

$$\sum_{k=1}^m \psi_k(t-h, t+h) < \varepsilon,$$

и неравенство (4) дает при достаточно малом $h > 0$

$$\Lambda(t+h) - \Lambda(t-h) < 2\varepsilon;$$

следовательно, $\Lambda(t+0) = \Lambda(t-0)$, и второе утверждение леммы 3 также доказано.

Лемма 4. *Финитный поток может иметь не более счетного множества сингулярных точек.*

Доказательство непосредственно вытекает из леммы 3.

Лемма 5. *Пусть L — финитный поток, и L' — поток тех событий потока L , моменты наступления которых отличны от некоторого фиксированного момента t_0 . Тогда t_0 есть регулярная точка потока L' .*

Доказательство. Пусть $0 < h_0 < t_0$. Обозначим через w_n вероятность того, что в области

$$\frac{h_0}{2^n} < |t - t_0| \leq \frac{h_0}{2^{n-1}} \quad (\Delta_n)$$

наступит по меньшей мере одно событие потока L' . Если промежуток $(t_0 - h_0, t_0 + h_0)$ содержит события потока L' , то то же самое должно иметь место и для по меньшей мере одной из областей (Δ_n) ($n=1, 2, \dots$). Поэтому, если обозначить вообще через $w(\alpha, \beta)$ вероятность наличия событий потока L' в промежутке (α, β) , мы будем иметь

$$w(t_0 - h_0, t_0 + h_0) \leq \sum_{n=1}^{\infty} w_n.$$

Аналогичным образом находим для любого натурального числа k

$$w\left(t_0 - \frac{h_0}{2^k}, t_0 + \frac{h_0}{2^k}\right) \leq \sum_{n=k+1}^{\infty} w_n. \quad (5)$$

Пусть теперь v_{nk} означает вероятность того, что в область (Δ_n) попадет k событий потока L' ; тогда

$$w_n = \sum_{k=1}^{\infty} v_{nk} \leq \sum_{k=1}^{\infty} k v_{nk} = \mu_n, \quad (6)$$

где μ_n означает математическое ожидание числа событий потока L' в области (Δ_n) . Так как поток L (а следовательно, и поток L') — финитный, то ряд $\sum_{n=1}^{\infty} \mu_n$ сходится; в силу (6)

сходится поэтому и ряд $\sum_{n=1}^{\infty} w_n$. Но тогда (5) показывает, что $w(t_0 - h, t_0 + h) \rightarrow 0$ ($h \rightarrow +0$), и, следовательно, t_0 есть регулярная точка потока L' , ч. т. д.

Условимся во всем дальнейшем называть финитный поток *регулярным*, если он не имеет ни одной сингулярной точки, так что любое t ($0 \leq t < +\infty$) является для него регулярной точкой. Тогда из леммы 3 непосредственно следует

Лемма 6. *Необходимым и достаточным условием регулярности потока служит непрерывность его ведущей функции.*

Наконец, нам понадобится еще следующая

Лемма 7. *Если $0 < \alpha < \beta$ и данный поток — регулярный, то равномерно в отрезке $\alpha \leq t \leq \beta$*

$$v_0(t-h, t+h) \rightarrow 1 \quad (h \rightarrow +0).$$

Доказательство. В силу леммы 1

$$\begin{aligned} 1 - v_0(t-h, t+h) &= \sum_{k=1}^{\infty} v_k(t-h, t+h) \leq \\ &\leq \sum_{k=1}^{\infty} k v_k(t-h, t+h) = \Lambda(t+h) - \Lambda(t-h), \end{aligned}$$

и утверждение леммы 7 вытекает из равномерной непрерывности функции $\Lambda(t)$ в любом конечном отрезке.

§ 3. Леммы о регулярных потоках без последействия

Лемма 8. *Для любого регулярного потока без последействия и любых α, β ($0 \leq \alpha < \beta$)*

$$v_0(\alpha, \beta) > 0.$$

Доказательство. Допустим, что $v_0(\alpha, \beta) = 0$, и положим $\gamma = \frac{1}{2}(\alpha + \beta)$. Так как данный поток — без последействия, то $v_0(\alpha, \beta) = v_0(\alpha, \gamma) v_0(\gamma, \beta)$ и, следовательно, либо $v_0(\alpha, \gamma) = 0$, либо $v_0(\gamma, \beta) = 0$. Безграничное продол-

жение этого процесса известным образом приводит нас к точке ξ , обладающей следующим свойством: при любом $h > 0$ промежуток $(\xi - h, \xi + h)$ содержит другой промежуток (a, b) , для которого $v_0(a, b) = 0$; но тогда и подавно при любом $h > 0$ мы имеем $v_0(\xi - h, \xi + h) = 0$, откуда $p_0(\xi) = 0$, что стоит в противоречии с регулярностью данного потока. Лемма 8 доказана.

Условимся во всем дальнейшем называть функцию

$$\Phi(x, \alpha, \beta) = \sum_{k=0}^{\infty} v_k(\alpha, \beta) x^k$$

(где ряд в правой части при любых α, β сходится по меньшей мере для $|x| \leq 1$) *производящей функцией* данного потока. Если этот поток — регулярный и без последствия, то по лемме 8 $v_0(\alpha, \beta) > 0$, и, следовательно, разложение

$$\ln \Phi(x, \alpha, \beta) = \sum_{k=0}^{\infty} \omega_k(\alpha, \beta) x^k$$

функции $\ln \Phi(x, \alpha, \beta)$ по степеням x сходится в некоторой окрестности точки $x = 0$. По известным свойствам производящих функций мы имеем для потоков без последствия при $\alpha < \gamma < \beta$

$$\Phi(x, \alpha, \beta) = \Phi(x, \alpha, \gamma) \Phi(x, \gamma, \beta),$$

и, следовательно,

$$\ln \Phi(x, \alpha, \beta) = \ln \Phi(x, \alpha, \gamma) + \ln \Phi(x, \gamma, \beta),$$

что в свою очередь имеет следствием

$$\omega_k(\alpha, \beta) = \omega_k(\alpha, \gamma) + \omega_k(\gamma, \beta) \quad (k \geq 0).$$

Поэтому, если положить $\omega_k(0, t) = \chi_k(t)$ ($t \geq 0$, $k \geq 0$), то $\omega_k(\alpha, \beta) = \chi_k(\beta) - \chi_k(\alpha)$, и мы получаем в некоторой окрестности точки $x = 0$ разложение

$$\ln \Phi(x, \alpha, \beta) = \sum_{k=0}^{\infty} [\chi_k(\beta) - \chi_k(\alpha)] x^k. \quad (7)$$

Для установления общего вида регулярного потока без последствия имеют существенное значение свойства функций $\chi_k(t)$; изучению этих свойств и будет посвящен конец настоящего параграфа.

Лемма 9. При $0 \leq \alpha < \beta$, $n \geq 1$ и достаточно малом $x > 0$

$$\frac{\partial^n \ln \Phi}{\partial x^n} = \frac{\Phi^{(n)}}{\Phi} + \sum a_{i_1 \dots i_n} \frac{\Phi^{(i_1)} \dots \Phi^{(i_n)}}{\Phi^n}; \quad (8)$$

здесь положено $\Phi = \Phi(x, \alpha, \beta)$, $\Phi^{(i)} = \frac{\partial^i \Phi(x, \alpha, \beta)}{\partial x^i}$; коэффициенты $a_{i_1 \dots i_n}$ зависят только от n и индексов $i_1, \dots, i_n$ (и не зависят от α, β и x); суммирование распространяется на все системы индексов $i_1, \dots, i_n$, подчиненные условиям

$$0 \leq i_k < n \quad (1 \leq k \leq n), \quad \sum_{k=1}^n i_k = n.$$

Соотношение (8) представляет собою элементарную формулу дифференциального исчисления, имеющую место для любой n -кратно дифференцируемой функции $\Phi(x)$ [$\Phi(0) = 0$]. Доказательство автоматически проводится индукцией по n [при $n=1$ формула (8) тривиальна] и может быть представлено читателю.

Лемма 10. Пусть $0 \leq \alpha < \beta$, $n \geq 1$ и данный поток — регулярный и без последствий. Пусть $\alpha = \alpha_0 < \alpha_1 < \dots < \alpha_s = \beta$, $d(s) = \max_{1 \leq r \leq s} (\alpha_r - \alpha_{r-1})$. Тогда в формуле (7)

$$\chi_n(\beta) - \chi_n(\alpha) = \lim_{d(s) \rightarrow 0} \sum_{r=1}^s \frac{v_n(\alpha_{r-1}, \alpha_r)}{v_0(\alpha_{r-1}, \alpha_r)}. \quad (9)$$

Предварительное замечание. В силу леммы 8 в правой части (9) все знаменатели отличны от нуля.

Доказательство. В ходе последующего рассуждения мы под $\Phi^{(i)}$ ($i=0, 1, \dots$) всегда будем понимать значение производной $\frac{\partial^i \Phi(x, \alpha, \beta)}{\partial x^i}$ при $x=0$.

Так как в силу (7)

$$\chi_n(\beta) - \chi_n(\alpha) = \frac{1}{n!} \left\{ \frac{\partial^n \ln \Phi(x, \alpha, \beta)}{\partial x^n} \right\}_{x=0} \quad (n \geq 0),$$

то соотношение (8) при $n \geq 1$ дает

$$\chi_n(\beta) - \chi_n(\alpha) = \frac{1}{n!} \left\{ \frac{\Phi^{(n)}}{\Phi} + \sum a_{i_1 \dots i_n} \frac{\Phi^{(i_1)} \dots \Phi^{(i_n)}}{\Phi^n} \right\}.$$

Так как, очевидно, $\Phi^{(i)} = i! v_i(\alpha, \beta)$ ($i \geq 0$), то отсюда

$$\chi_n(\beta) - \chi_n(\alpha) = \frac{v_n(\alpha, \beta)}{v_0(\alpha, \beta)} + \sum b_{i_1 \dots i_n} \frac{v_{i_1}(\alpha, \beta) \dots v_{i_n}(\alpha, \beta)}{[v_0(\alpha, \beta)]^n}, \quad (10)$$

где область суммирования определяется соотношениями

$$0 \leq i_k < n \quad (1 \leq k \leq n), \quad \sum_{k=1}^n i_k = n,$$

а коэффициенты $b_{i_1 \dots i_n}$, подобно прежним $a_{i_1 \dots i_n}$, не зависят от α и β . Если мы разобьем промежуток (α, β) на s промежутков (α_{r-1}, α_r) ($1 \leq r \leq s$), как это описано в формулировке леммы 10, то соотношение (10) имеет место для всех промежутков (α_{r-1}, α_r) , и притом с одними и теми же коэффициентами $b_{i_1 \dots i_n}$. Написав все эти соотношения и складывая их почленно, мы получаем

$$\begin{aligned} \chi_n(\beta) - \chi_n(\alpha) &= \sum_{r=1}^s \frac{v_n(\alpha_{r-1}, \alpha_r)}{v_0(\alpha_{r-1}, \alpha_r)} + \\ &+ \sum b_{i_1 \dots i_n} \sum_{r=1}^s \frac{v_{i_1}(\alpha_{r-1}, \alpha_r) \dots v_{i_n}(\alpha_{r-1}, \alpha_r)}{[v_0(\alpha_{r-1}, \alpha_r)]^n}. \end{aligned} \quad (11)$$

Чтобы доказать лемму 10, нам надо только установить, что при $d(s) = \max_{1 \leq r \leq s} (\alpha_r - \alpha_{r-1}) \rightarrow 0$ второе слагаемое правой части формулы (11) становится бесконечно малым. Это же в свою очередь будет доказано, если мы убедимся, что для любой фиксированной, принадлежащей области суммирования системы индексов $i_1, i_2, \dots, i_n$ величина

$$S = \sum_{r=1}^s \frac{v_{i_1}(\alpha_{r-1}, \alpha_r)}{v_0(\alpha_{r-1}, \alpha_r)} \frac{v_{i_2}(\alpha_{r-1}, \alpha_r)}{v_0(\alpha_{r-1}, \alpha_r)} \dots \frac{v_{i_n}(\alpha_{r-1}, \alpha_r)}{v_0(\alpha_{r-1}, \alpha_r)}$$

стремится к нулю при $d(s) \rightarrow 0$.

С этой целью заметим прежде всего, что при $n=1$ область суммирования $0 \leq i_k < n$ ($1 \leq k \leq n$), $\sum_{k=1}^n i_k = n$ ста-

новится пустой (второе слагаемое правой части (11)). Поэтому мы допустим, что $n > 1$ и что для всех чисел $< n$ лемма 10 уже доказана.

Обратимся теперь к исследованию величины S . В силу леммы 7 при любом $\varepsilon > 0$ и достаточно малом $d(s)$

$$1 - v_0(\alpha_{r-1}, \alpha_r) < \varepsilon \quad (1 \leq r \leq s),$$

и, следовательно,

$$\frac{1 - v_0(\alpha_{r-1}, \alpha_r)}{v_0(\alpha_{r-1}, \alpha_r)} < \frac{\varepsilon}{1 - \varepsilon} \quad (1 \leq r \leq s).$$

Но при $i_k > 0$ мы имеем

$$v_{i_k}(\alpha_{r-1}, \alpha_r) \leq 1 - v_0(\alpha_{r-1}, \alpha_r),$$

и, следовательно,

$$\frac{v_{i_k}(\alpha_{r-1}, \alpha_r)}{v_0(\alpha_{r-1}, \alpha_r)} < \frac{\varepsilon}{1 - \varepsilon} \quad (1 \leq r \leq s),$$

в то время как при $i_k = 0$ это отношение равно 1. Но в силу условий $i_k < n$ ($1 \leq k \leq n$) и $\sum_{k=1}^n i_k = n$ среди индексов i_k ($1 \leq k \leq n$) найдется по крайней мере два положительных. Пусть, например, такими будут индексы i и l . Тогда приведенные выше оценки дают

$$S \leq \sum_{r=1}^s \frac{v_i(\alpha_{r-1}, \alpha_r)}{v_0(\alpha_{r-1}, \alpha_r)} \frac{v_l(\alpha_{r-1}, \alpha_r)}{v_0(\alpha_{r-1}, \alpha_r)} < \frac{\varepsilon}{1 - \varepsilon} \sum_{r=1}^s \frac{v_l(\alpha_{r-1}, \alpha_r)}{v_0(\alpha_{r-1}, \alpha_r)}.$$

Так как $l < n$, то мы можем принять, что лемма 10 уже доказана с l вместо n ; поэтому последняя сумма при $d(s) \rightarrow 0$ имеет пределом $\chi_l(\beta) - \chi_l(\alpha)$. А так как ε при достаточно малом $d(s)$ может быть выбрано сколь угодно малым, то $S \rightarrow 0$ при $d(s) \rightarrow 0$, чем лемма 10 и доказана.

Лемма 11. При $n \geq 1$, $\chi_n(t)$ есть неубывающая функция от t ; ряд

$$\sum_{n=0}^{\infty} n [\chi_n(\beta) - \chi_n(\alpha)]$$

сходится при любых α, β ; для всех $t \geq 0$

$$\sum_{n=0}^{\infty} \chi_n(t) = 0.$$

Доказательство. Первое утверждение непосредственно вытекает из леммы 10, в силу которой $\chi_n(\beta) - \chi_n(\alpha) \geq 0$ ($\beta \geq \alpha$). Далее, из леммы 10 следует при любом $r \geq 1$

$$\begin{aligned} \sum_{l=1}^m l[\chi_l(\beta) - \chi_l(\alpha)] &= \lim_{d(s) \rightarrow 0} \sum_{r=1}^s \frac{1}{v_0(\alpha_{r-1}, \alpha_r)} \sum_{l=1}^m l v_l(\alpha_{r-1}, \alpha_r) \leq \\ &\leq \lim_{d(s) \rightarrow 0} \sup \sum_{r=1}^s \frac{\Lambda(\alpha_r) - \Lambda(\alpha_{r-1})}{v_0(\alpha_{r-1}, \alpha_r)} = \Lambda(\beta) - \Lambda(\alpha) \end{aligned}$$

[последнее равенство вытекает из того, что в силу леммы 7 $v_0(\alpha_{r-1}, \alpha_r)$ при $d(s) \rightarrow 0$ стремится к 1 равномерно относительно r ($1 \leq r \leq s$)]. Так как правая часть этого равенства не зависит от m , то доказано и второе утверждение леммы. Наконец, из этого второго утверждения следует, что формула (7) имеет место при $|x| \leq 1$. В частности, при $x = 1$ это дает

$$0 = \ln \Phi(1, \alpha, \beta) = \sum_{k=0}^{\infty} [\chi_k(\beta) - \chi_k(\alpha)],$$

чем доказано и третье утверждение леммы 11.

Лемма 12. Каждая функция $\chi_k(t)$ ($k = 0, 1, \dots$) всюду непрерывна.

Доказательство. Допустим, что для некоторого $k \geq 1$ и некоторого $t \geq 0$

$$\chi_k(t+0) - \chi_k(t-0) > 0.$$

Так как все функции $\chi_k(t)$ ($k \geq 1$) — неубывающие, то тогда

$$\chi_0(t+0) - \chi_0(t-0) = - \sum_{k=1}^{\infty} [\chi_k(t+0) - \chi_k(t-0)] < 0.$$

Но

$$\ln \Phi(0, t-h, t+h) = \chi_0(t+h) - \chi_0(t-h).$$

Поэтому при достаточно малом $h > 0$

$$\ln \Phi(0, t-h, t+h) < -\alpha,$$

или

$$\Phi(0, t-h, t+h) = v_0(t-h, t+h) < e^{-\alpha},$$

где α — положительная постоянная. А отсюда следует, что

$$\lim_{h \rightarrow +0} v_0(t-h, t+h) \leq e^{-\alpha} < 1,$$

что противоречит допущению о регулярности данного потока. Этим лемма 12 доказана.

§ 4. Общая форма регулярных потоков без последействия

Леммы § 3 в своей совокупности показывают, что производящая функция $\Phi(x, \alpha, \beta)$ любого регулярного потока без последействия имеет вид

$$\Phi(x, \alpha, \beta) = \exp \left\{ \sum_{k=0}^{\infty} [\chi_k(\beta) - \chi_k(\alpha)] x^k \right\}, \quad (12)$$

где функции $\chi_k(t)$ ($\chi_k(0) = 0$) обладают следующими свойствами:

1° Все функции $\chi_k(t)$ всюду непрерывны.
 2° При любом k функция $\chi_k(t)$ монотонна (неубывающая при $k > 0$ и невозрастающая при $k = 0$).

3° Ряд $\sum_{k=0}^{\infty} k \chi_k(t)$ сходится при любом $t \geq 0$.

4° Сумма $\sum_{k=0}^{\infty} \chi_k(t) = 0$ при любом $t \geq 0$.

Теперь мы докажем обратное предложение: если $\Phi(x, \alpha, \beta)$ есть функция вида (12), где функции $\chi_k(t)$ удовлетворяют требованиям 1°—4°, то существует регулярный поток без последействия, производящая функция которого совпадает с функцией $\Phi(x, \alpha, \beta)$.

Так как в силу свойства 2° разность $\chi_k(\beta) - \chi_k(\alpha)$ при $k > 0$ и $\alpha < \beta$ неотрицательна, то, как известно, существует поток без последействия L_k , для которого

$$v_m^{(k)}(\alpha, \beta) = \begin{cases} \exp \{ - [\chi_k(\beta) - \chi_k(\alpha)] \} \frac{[\chi_k(\beta) - \chi_k(\alpha)]^r}{r!} & (m = kr) \\ 0 & (m \not\equiv 0 \pmod k) \end{cases}$$

$[v_m^{(k)}(\alpha, \beta)]$ означает вероятность того, что в промежутке (α, β) наступит m событий потока L_k ; требуемое для этого условие, чтобы при $\alpha < \gamma < \beta$ всего было

$$v_m^{(k)}(\alpha, \beta) = \sum_{i=0}^m v_i^{(k)}(\alpha, \gamma) v_{m-i}^{(k)}(\gamma, \beta),$$

легко проверяется непосредственным подсчетом. Производящая функция потока L_k равна

$$\begin{aligned} \Phi_k(x, \alpha, \beta) &= \sum_{r=0}^{\infty} v_{rk}^{(k)}(\alpha, \beta) x^{rk} = \\ &= \exp \{ - [\chi_k(\beta) - \chi_k(\alpha)] \} \sum_{r=0}^{\infty} \frac{[\chi_k(\beta) - \chi_k(\alpha)]^r}{r!} x^{rk} = \\ &= \exp \{ [\chi_k(\beta) - \chi_k(\alpha)] (x^k - 1) \}. \end{aligned}$$

Рассмотрим теперь суперпозицию L определенных нами потоков $L_1, L_2, \dots, L_n, \dots$, которые будем предполагать взаимно независимыми. Легко убедиться, что производящая функция потока L (который, очевидно, есть поток без последдействия) совпадает с

$$F(x, \alpha, \beta) = \prod_{k=1}^{\infty} \Phi_k(x, \alpha, \beta).$$

В самом деле, если положить

$$F(x, \alpha, \beta) = \sum_{r=0}^{\infty} v_r(\alpha, \beta) x^r$$

(этим соотношением определяется число $v_r(\alpha, \beta)$), то, очевидно, для любого $r \geq 0$

$$v_r(\alpha, \beta) = \sum v_{r_1}^{(1)}(\alpha, \beta) v_{r_2}^{(2)}(\alpha, \beta) \dots v_{r_k}^{(k)}(\alpha, \beta) \dots, \quad (13)$$

где суммирование распространяется на область

$$r_k \geq 0 \quad (k \geq 1), \quad \sum_{k=1}^{\infty} r_k = r.$$

Но в силу взаимной независимости потоков L_k сумма в правой части (13), очевидно, равна вероятности того, что

в промежутке (α, β) наступит r событий суммарного потока L . В силу (13), число $v_r(\alpha, \beta)$ совпадает с этой вероятностью, а это и значит, что $F(x, \alpha, \beta)$ есть производящая функция потока L .

Но в силу свойства 4° мы имеем

$$\begin{aligned} F(x, \alpha, \beta) &= \prod_{k=1}^{\infty} \Phi_k(x, \alpha, \beta) = \\ &= \exp \left\{ \sum_{k=1}^{\infty} [\chi_k(\beta) - \chi_k(\alpha)] (x^k - 1) \right\} = \\ &= \exp \left\{ \sum_{k=1}^{\infty} [\chi_k(\beta) - \chi_k(\alpha)] x^k - \sum_{k=1}^{\infty} [\chi_k(\beta) - \chi_k(\alpha)] \right\} = \\ &= \exp \left\{ \sum_{k=0}^{\infty} [\chi_k(\beta) - \chi_k(\alpha)] x^k \right\} = \Phi(x, \alpha, \beta). \end{aligned}$$

Этим показано, что данная функция $\Phi(x, \alpha, \beta)$ действительно есть производящая функция некоторого потока без последствия L . Нам остается только убедиться, что построенный нами поток L — регулярный.

В силу свойства 3° для любых $\varepsilon > 0$ и $t > 0$ найдется такое $N > 0$, что

$$\sum_{k > N} k \chi_k(t) < \varepsilon. \quad (14)$$

Так как поток $L_k (k=1, 2, \dots)$ имеет непрерывную (в силу свойства 1°) ведущую функцию $k \chi_k(t)$, то он регулярен в силу леммы 6. Следовательно, для любой фиксированной точки $\gamma (0 < \gamma < t)$ при достаточно малом $h_0 > 0$ и при $0 < h < h_0$

$$\sum_{k=1}^N [1 - v_0^{(k)}(\gamma - h, \gamma + h)] < \varepsilon. \quad (15)$$

Но при $0 \leq \alpha < \beta$

$$\begin{aligned} 1 - v_0^{(k)}(\alpha, \beta) &= \sum_{p=1}^{\infty} v_p^{(k)}(\alpha, \beta) \leq \sum_{p=1}^{\infty} p v_p^{(k)}(\alpha, \beta) = \\ &= k [\chi_k(\beta) - \chi_k(\alpha)]. \end{aligned}$$

[так как $k\chi_k(t)$ есть ведущая функция потока L_k]. Из (15) и (14) следует поэтому при $0 < h < h_0$, $0 < \gamma - h < \gamma + h < t$

$$\sum_{k=1}^{\infty} [1 - v_0^{(k)}(\gamma - h, \gamma + h)] \leq \sum_{k=1}^N [1 - v_0^{(k)}(\gamma - h, \gamma + h)] + \\ + \sum_{k>N} [1 - v_0^{(k)}(0, t)] < \varepsilon + \sum_{k>N} k\chi_k(t) < 2\varepsilon.$$

Поэтому для суммарного потока L мы получаем при $0 < h < h_0$

$$1 - v_0(\gamma - h, \gamma + h) \leq \sum_{k=1}^{\infty} [1 - v_0^{(k)}(\gamma - h, \gamma + h)] < 2\varepsilon$$

$[1 - v_0(\gamma - h, \gamma + h)]$ есть вероятность того, что в промежутке $(\gamma - h, \gamma + h)$ наступит по меньшей мере одно событие потока L . Но это означает, что γ — регулярная точка потока L ; а так как выбор этой точки произволен, то поток L — регулярный, ч. т. д.

Мы можем, таким образом, сформулировать результаты §§ 3 и 4 в виде следующего предложения.

Теорема I. *Каждый регулярный поток без последствия имеет производящую функцию $\Phi(x, \alpha, \beta)$ вида (12), где функции $\chi_k(t)$ удовлетворяют требованиям 1°—4°; обратно, каждая такая функция $\Phi(x, \alpha, \beta)$ есть производящая функция некоторого регулярного потока без последствия.*

Посмотрим теперь еще, какой вид принимает общая форма (12) для некоторых известных классов потоков без последствия.

1. Пусть данный поток при любом $\alpha \geq 0$ удовлетворяет требованию

$$1 - v_0(\alpha, \beta) - v_1(\alpha, \beta) = \sum_{k=2}^{\infty} v_k(\alpha, \beta) = o(\beta - \alpha) \\ (\beta \rightarrow \alpha + 0)$$

ср. [1], § 1), и пусть при любом $\alpha \geq 0$ существует предел

$$\lim_{\beta \rightarrow \alpha + 0} \frac{1 - v_0(\alpha, \beta)}{\beta - \alpha} = \lambda(\alpha).$$

Тогда формула (9) леммы 10 легко дает

$$\chi_1(\beta) - \chi_1(\alpha) = \int_{\alpha}^{\beta} \lambda(u) du, \quad \chi_k(\beta) - \chi_k(\alpha) = 0 \quad (k > 1),$$

откуда в силу свойства 4°

$$\chi_0(\beta) - \chi_0(\alpha) = - \int_{\alpha}^{\beta} \lambda(u) du,$$

и формула (12) получает вид

$$\Phi(x, \alpha, \beta) = \exp \left\{ (x-1) \int_{\alpha}^{\beta} \lambda(u) du \right\}.$$

Это — давно известный результат (см. [1], стр. 17*), или [6], стр. 183).

2. Пусть данный поток — стационарный; положим

$$\lim_{t \rightarrow +0} \frac{1 - v_0(t)}{t} = \lambda, \quad \lim_{t \rightarrow +0} \frac{v_k(t)}{t} = \lambda p_k \quad (k > 0);$$

тогда формула (9) леммы 10 дает при любом $t \geq 0$

$$\chi_k(t) = \lambda t p_k \quad (k \geq 1), \quad \chi_0(t) = -\lambda t,$$

и формула (12) получает вид

$$\Phi(x, 0, t) = \exp \left\{ \lambda t \left(\sum_{k=1}^{\infty} p_k x^k - 1 \right) \right\}.$$

Это полностью совпадает с формулой (8.6) ([1], стр. 31) и является известным результатом Редхеффера [2] и автора.

§ 5. Общая форма финитных потоков без последействия

До сих пор мы рассматривали только регулярные потоки, т. е. такие, ведущая функция $\Lambda(t)$ которых всюду непрерывна. Теперь мы обратимся к рассмотрению любого финитного потока без последействия, так что функция $\Lambda(t)$ всюду

*) Пользуюсь случаем указать на грубую опечатку: под знаком последней суммы стр. 17 [1] пропущен множитель x^k .

конечна, но может быть как угодно разрывной. Как мы знаем из § 2, точки разрыва функции $\Lambda(t)$ — это сингулярные точки данного потока.

Условимся теперь называть поток S *сингулярным*, если он обладает следующими тремя свойствами:

1) События потока S могут наступать лишь в некоторые заранее определенные моменты времени; эти моменты, которые мы будем называть *ступенями* потока S , образуют не более чем счетное множество.

2) Числа событий, падающих на различные ступени, представляют собою взаимно независимые случайные величины.

3) Пусть $q_k^{(i)}$ ($k \geq 0$, $i \geq 1$) означает вероятность того, что на ступень t_i придется k событий; тогда при любом $t > 0$

$$\sum_{0 \leq t_i \leq t} \sum_{k=0}^{\infty} k q_k^{(i)} < +\infty.$$

Свойство 2) выражает собой отсутствие последействия, а свойство 3) — финитность потока S . Очевидно, сингулярный поток однозначно определяется заданием чисел t_i и $q_k^{(i)}$ ($i \geq 1$, $k \geq 0$). Структура такого потока представляется поэтому совершенно прозрачной: выбор чисел t_i и $q_k^{(i)}$ подчинен только требованиям $t_i \geq 0$, $q_k^{(i)} \geq 0$ и $\sum_{k=0}^{\infty} q_k^{(i)} = 1$ ($i = 1, 2, \dots$), в остальном же остается произвольным.

Пусть теперь нам дан произвольный финитный поток K без последействия. Обозначим через $t_1, t_2, \dots$ сингулярные точки этого потока, занумерованные в произвольном порядке. Мы будем рассматривать поток K как суперпозицию двух потоков R и S . Будем считать некоторое событие потока K принадлежащим потоку S , если оно наступает в один из моментов $t_1, t_2, \dots$, в противном случае мы будем считать его принадлежащим потоку R . Можно убедиться (чего мы здесь делать не будем), что оба потока S и R — без последействия, и что они взаимно независимы (ср. работу Марчевского [5]).

Докажем в первую очередь регулярность потока R . Пусть K_i ($i \geq 1$) означает совокупность событий потока K , моменты наступления которых отличны от t_i . В силу леммы 5 точка t_i является регулярной для потока K_i , а значит и

подавно — для потока $R \subset K_i$. Но точка t_i , отличная от всех t_l , по самому определению чисел t_l регулярна для потока K , а значит и для $R \subset K$. Таким образом, поток R не имеет сингулярных точек и, следовательно, регулярен.

Что касается потока S , то сингулярность его очевидна из самого определения; ступенями его служат числа t_i ($i \geq 1$). Мы покажем еще, что его характеристические вероятности $q_k^{(i)}$ соответственно совпадают с определенными в § 2 числами

$$p_k(t_i) = \lim_{h \rightarrow +0} v_k(t_i - h, t_i + h) \quad (k \geq 0, i \geq 1).$$

Пусть поток K_i ($i \geq 1$) определяется как выше, и пусть $v_k^{(i)}(\alpha, \beta)$ означает вероятность того, что в промежутке (α, β) наступит k событий потока K_i . Тогда для потока K мы, очевидно, имеем при $i \geq 1, k \geq 0$

$$v_k(t_i - h, t_i + h) = \sum_{l=0}^k q_l^{(i)} v_{k-l}^{(i)}(t_i - h, t_i + h). \quad (16)$$

Так как t_i (в силу леммы 5) есть регулярная точка потока K_i , то при $h \rightarrow +0$

$$v_{k-l}^{(i)}(t_i - h, t_i + h) \rightarrow \begin{cases} 1 & (l=k), \\ 0 & (l < k). \end{cases}$$

Поэтому соотношение (16) при $h \rightarrow +0$ дает в пределе

$$p_k(t_i) = \lim_{h \rightarrow +0} v_k(t_i - h, t_i + h) = q_k^{(i)} \quad (i \geq 1, k \geq 0),$$

ч. т. д.

Резюмируя результаты настоящего параграфа, мы приходим к следующему предложению.

Теорема II. *Всякий финитный поток без последействия может рассматриваться как суперпозиция двух взаимно независимых потоков того же типа, из которых один — регулярный, а другой — сингулярный; ступенями t_i сингулярной компоненты служат при этом сингулярные точки данного потока, а характеристическими вероятностями $q_k^{(i)}$ — соответствующие числа $p_k(t_i)$ ($i \geq 1, k \geq 0$) данного потока.*

Непосредственно очевидно, что и обратно — любая суперпозиция описанного типа представляет собою некоторый финитный поток без последействия.

Теоремы I и II вместе взятые дают легко обозримое описание совокупности всех финитных потоков случайных событий без последействия, что и было целью настоящего исследования.

ЛИТЕРАТУРА

- [1] А. Я. Хинчин, Математические методы теории массового обслуживания, Труды Матем. инст. им. Стеклова, 49, (1955), 1—122.
 - [2] R. M. Redheffer, Math. Magazine, 26, (1953), 185—188.
 - [3] C. Palm, Intensitätsschwankungen im Fernspreckverkehr, Ericsson Technics, 44 (1943), 1—189.
 - [4] L. Takacs, Acta math. hung., 5, (1954), 203—236.
 - [5] D. Marzewski, Studia math. (Warszawa), 13, No 1, (1953), 130—136.
 - [6] Фрай Торнтон, Теория вероятностей для инженеров М. —Л., ГТТИ, 1934.
-

О ПУАССОНОВСКИХ ПОТОКАХ СЛУЧАЙНЫХ СОБЫТИЙ*)

§ 1. Постановка задачи

В классических приложениях теории потоков случайных событий (например, в теории распада атомов и теории массового обслуживания) рассматривается, как известно, в большинстве случаев лишь простейший вид таких потоков, когда вероятность наступления k событий в произвольно взятом промежутке времени длины t дается формулой

$$v_k(t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}, \quad (1)$$

где $\lambda > 0$ — постоянная; такой поток предполагается обычно потоком *без последствия*; это означает, что для любой конечной подгруппы попарно не пересекающихся промежутков времени числа наступающих в этих промежутках событий представляют собой взаимно независимые случайные величины.

Причиной столь широкой распространенности потоков специальной формы (1) служит, как известно, то, что эта форма для некоторого весьма обширного класса потоков фактически является единственно возможной. Именно, если к требованиям стационарности и отсутствия последствия добавить еще условие

$$\psi_2(t) = o(t) \quad (t \rightarrow 0), \quad (2)$$

где $\psi_2(t)$ означает вероятность наступления по меньшей мере двух событий в промежутке времени длины t , то рассматриваемый поток обязательно будет иметь форму (1)

*) Теория вероятностей и ее применения, т. 1, вып. 3, 1956, стр. 320—327.

(см., например, [1], гл. 1). Обратно, всякий поток без последствия, имеющий форму (1), очевидно стационарен и, как легко убедиться, удовлетворяет условию (2). Таким образом, три перечисленных свойства — отсутствие последствия, стационарность и условие (2) — полностью характеризуют собою класс «пуассоновских» потоков (1).

Однако в приложениях все чаще появляются задачи, требующие исследования *нестационарных* потоков событий без последствия. За исходный пункт такого исследования обычно бывает целесообразно принять «ведущую функцию» $\Lambda(t)$ данного потока — математическое ожидание числа событий, наступающих в полуоткрытом промежутке $[0, t)$. В стационарном случае $\Lambda(t) = \lambda t$, где $\lambda > 0$ — постоянная; в общем случае $\Lambda(t)$ — любая неубывающая функция от t [$\Lambda(t) = 0$ при $t \leq 0$]. Обычно при этом предполагают, что поток имеет «пуассоновскую» форму

$$v_k(\alpha, \beta) = e^{-[\Lambda(\beta) - \Lambda(\alpha)]} \frac{[\Lambda(\beta) - \Lambda(\alpha)]^k}{k!} \quad (3)$$

$[v_k(\alpha, \beta)]$ означает вероятность того, что в промежутке (α, β) наступит k событий данного потока; задание всех функций $v_k(\alpha, \beta)$ ($k \geq 0$, $0 \leq \alpha < \beta$), очевидно, дает полное описание данного потока без последствия]. Однако форма (3) при заданной ведущей функции $\Lambda(t)$ отнюдь не является единственно возможной: ведь даже в стационарном случае, как было упомянуто выше, форма (3) [в этом случае совпадающая с (1)] имеет место лишь при выполнении дополнительного требования (2). С другой стороны, в моей недавней работе [2] непосредственно показано, что общая форма потоков без последствия с данной ведущей функцией $\Lambda(t)$, даже в случае непрерывной $\Lambda(t)$ выходит далеко за пределы формулы (3).

Таким образом, естественно возникает вопрос о том, какими свойствами должен обладать поток без последствия для того, чтобы иметь простейшую (почти всегда принимаемую в приложениях) «пуассоновскую» форму (3); дело идет при этом об отыскании возможно простого (необходимого и достаточного) дополнительного требования, подобно условию (2) в стационарном случае. Решение этой задачи означало бы распространение на нестационарный случай той фундаментальной теоремы, которую мы приводили в самом

начале для случая стационарного. От данного потока при этом, кроме отсутствия последствия, мы будем требовать только существование ведущей функции $\Lambda(t)$ [т. е. $\Lambda(t) < +\infty$ при $t < +\infty$].

Как покажет исследование, наиболее интересным при этом будет случай, когда функция $\Lambda(t)$ непрерывна; точки же разрыва этой функции, если они имеются, осложняют решение задачи лишь в весьма незначительной степени.

§ 2. Регулярный случай

Условимся ради краткости называть поток без последствия регулярным, если его ведущая функция $\Lambda(t)$ всюду непрерывна. В этом параграфе мы будем рассматривать только регулярные потоки. В частности, регулярным будет любой стационарный поток без последствия [так как для такого потока $\Lambda(t) = \lambda t$, $0 < \lambda < +\infty$].

Пусть для любого стационарного потока и любого $k \geq 0$, $\psi_k(t)$ означает вероятность того, что в промежутке длины t наступит по меньшей мере k событий данного потока. Очевидно, мы всегда имеем

$$\psi_k(t) = \sum_{l=k}^{\infty} v_l(t) \quad (k \geq 0).$$

Убедимся прежде всего, что для стационарного потока условие (2) равносильно соотношению

$$\frac{\psi_2(t)}{\psi_1(t)} \rightarrow 0 \quad (t \rightarrow 0). \quad (4)$$

С этой целью заметим, что, как известно ([1], стр. 24), для любого стационарного потока при $t \rightarrow +0$

$$\frac{\psi_1(t)}{t} \rightarrow \alpha \quad (0 < \alpha \leq +\infty);$$

но в нашем случае

$$\psi_1(t) = \sum_{k=1}^{\infty} v_k(t) \leq \sum_{k=1}^{\infty} k v_k(t) = \Lambda(t) = \lambda t$$

и, следовательно, $\alpha \leq \lambda < +\infty$; таким образом, при $t \rightarrow +0$ отношение $\psi_1(t)/t$ в нашем случае стремится к некоторому

конечному положительному пределу, и соотношение

$$\frac{\psi_2(t)}{t} = \frac{\psi_2(t)}{\psi_1(t)} \frac{\psi_1(t)}{t}$$

показывает, что требования (2) и (4) действительно равносильны.

Условимся теперь называть любой (вообще говоря, нестационарный) поток *ординарным*, если при любых $t > 0$ и $\varepsilon > 0$ найдется такое $\delta > 0$, что для $0 \leq \alpha < \beta < \alpha + \delta < t$ имеет место неравенство

$$\psi_2(\alpha, \beta) \leq \varepsilon \psi_1(\alpha, \beta). \quad (5)$$

При этом $\psi_k(\alpha, \beta)$ означает вероятность того, что в промежутке (α, β) наступит по меньшей мере k событий данного потока. Вместо этого несколько сложного определения ординарности мы могли бы и просто сказать, что

$$\frac{\psi_2(\alpha, \beta)}{\psi_1(\alpha, \beta)} \rightarrow 0 \quad (\beta - \alpha \rightarrow 0)$$

равномерно в области $0 \leq \alpha < \beta \leq t$; если бы мы не были вынуждены учитывать возможность равенства $\psi_1(\alpha, \beta) = 0$.

Для стационарного потока ординарность, очевидно, равносильна требованию (4) (а значит, как мы видели, и требованию (2)). Приведенная нами в § 1 хорошо известная теорема может поэтому быть сформулирована так, что для стационарного потока без последствий ординарность служит *необходимой и достаточной предпосылкой формы (1)*. Теперь мы покажем, что эта теорема распространяется на все регулярные потоки: *необходимой и достаточной предпосылкой пуассоновской формы (3) регулярного потока является его [понимаемая в смысле условия (5)] ординарность*.

Доказательство. Пусть данный поток — регулярный и ординарный. Его производящая функция

$$\Phi(x, \alpha, \beta) = \sum_{k=0}^{\infty} v_k(\alpha, \beta) x^k,$$

как я показал в моей работе [2], может быть представлена в виде

$$\Phi(x, \alpha, \beta) = \exp \left\{ \sum_{k=0}^{\infty} [\chi_k(\beta) - \chi_k(\alpha)] x^k \right\},$$

где функции $\chi_k(t)$ ($k=0, 1, \dots$) всюду непрерывны и при $k > 0$ — не убывающие; при этом

$$\chi_k(0) = 0 \quad (k \geq 0), \quad \sum_{k=0}^{\infty} \chi_k(t) = 0, \quad \sum_{k=1}^{\infty} k \chi_k(t) < +\infty \quad (t \geq 0).$$

Там же мною показано, что при $k > 0$

$$\chi_k(\beta) - \chi_k(\alpha) = \lim_{d \rightarrow 0} \sum_{r=1}^s \frac{v_k(\alpha_{r-1}, \alpha_r)}{v_0(\alpha_{r-1}, \alpha_r)}, \quad (6)$$

где $\alpha = \alpha_0 < \alpha_1 < \dots < \alpha_s = \beta$ — любое разбиение промежутка (α, β) , и $d = \max_{1 \leq r \leq s} (\alpha_r - \alpha_{r-1})$.

В нашем случае, в силу предположенной ординарности данного потока, для любых $t > 0$ и $\varepsilon > 0$ найдется такое $\delta > 0$, что при $0 \leq t_1 < t_2 < t_1 + \delta < t$ всегда

$$\psi_2(t_1, t_2) \leq \varepsilon \psi_1(t_1, t_2).$$

Отсюда

$$(1 - \varepsilon) \psi_2(t_1, t_2) \leq \varepsilon [\psi_1(t_1, t_2) - \psi_2(t_1, t_2)] = \varepsilon v_1(t_1, t_2)$$

и, следовательно, при $\varepsilon < 1/2$

$$\psi_2(t_1, t_2) \leq 2\varepsilon v_1(t_1, t_2).$$

Поэтому, если δ достаточно мало, мы имеем при $d < \delta$

$$\psi_2(\alpha_{r-1}, \alpha_r) \leq 2\varepsilon v_1(\alpha_{r-1}, \alpha_r) \quad (1 \leq r \leq s),$$

и при $k > 1$ из (6) следует

$$\begin{aligned} \chi_k(\beta) - \chi_k(\alpha) &\leq \limsup_{d \rightarrow 0} \sum_{r=1}^s \frac{\psi_k(\alpha_{r-1}, \alpha_r)}{v_0(\alpha_{r-1}, \alpha_r)} \leq \\ &\leq \limsup_{d \rightarrow 0} \sum_{r=1}^s \frac{\psi_2(\alpha_{r-1}, \alpha_r)}{v_0(\alpha_{r-1}, \alpha_r)} \leq 2\varepsilon \limsup_{d \rightarrow 0} \sum_{r=1}^s \frac{v_1(\alpha_{r-1}, \alpha_r)}{v_0(\alpha_{r-1}, \alpha_r)} = \\ &= 2\varepsilon [\chi_1(\beta) - \chi_1(\alpha)]. \end{aligned}$$

А так как ε может быть выбрано сколь угодно малым, то $\chi_k(\beta) - \chi_k(\alpha) = 0$ ($k > 1$), и мы находим

$$\Phi(x, \alpha, \beta) = \exp \{ [\chi_0(\beta) - \chi_0(\alpha)] + x [\chi_1(\beta) - \chi_1(\alpha)] \}.$$

или, полагая $\chi_1(t) = -\chi_0(t) = \Lambda(t)$,

$$\Phi(x, \alpha, \beta) = \exp \{ [\Lambda(\beta) - \Lambda(\alpha)](x - 1) \},$$

что равносильно (3).

Пусть мы теперь имеем любой регулярный поток вида (3). Полагая $\Lambda(\beta) - \Lambda(\alpha) = h$, мы находим

$$\psi_1(\alpha, \beta) = 1 - e^{-h}, \quad \psi_2(\alpha, \beta) = 1 - e^{-h} - he^{-h}; \quad (7)$$

так как функция $\Lambda(t)$ равномерно непрерывна в любом конечном интервале, то величина h при $\beta - \alpha \rightarrow 0$ бесконечно мала равномерно относительно α и β . Неравенство (5) тривиально при $h=0$; если же $h > 0$, то в силу (7) $\psi_1(\alpha, \beta) = h + o(h)$, $\psi_2(\alpha, \beta) = o(h)$, откуда снова вытекает (5), если $\beta - \alpha$ (а следовательно, и h) достаточно мало. Таким образом, данный поток — ординарный, и наше утверждение доказано полностью.

§ 3. Общий случай

Переходим теперь к случаю, когда $\Lambda(t)$ — любая неубывающая функция [$\Lambda(t) = 0$ при $t < 0$]. В моей уже цитированной работе [2] я называл данный поток сингулярным, если образующие его события могут наступать лишь в заранее определенные моменты времени t_i ($i = 1, 2, \dots$) (в конечном или счетном числе), а числа событий, наступающих в различные моменты t_i , представляют собою взаимно независимые случайные величины. В частности, если вероятность наступления k событий в момент t_i имеет вид

$$e^{-\alpha_i} \frac{\alpha_i^k}{k!} \quad (\alpha_i > 0, \quad i = 1, 2, \dots), \quad (8)$$

т. е. если каждая из только что упомянутых случайных величин распределена по некоторому закону Пуассона, то мы будем называть данный сингулярный поток пуассоновским. Для полного описания такого потока достаточно, таким образом, задать все его «ступени» t_i и соответствующие им значения α_i пуассоновского параметра.

Если имеется только одна ступень t_i и число наступающих в момент t_i событий подчиняется пуассоновскому закону (8), то производящая функция числа событий, наступающих

в промежутке (α, β) , равна

$$\Gamma_i(x, \alpha, \beta) = \begin{cases} \exp\{\alpha_i(x-1)\} & (\alpha \leq t_i < \beta), \\ 1 & \text{в противном случае} \end{cases}$$

Но в общем случае пуассоновский сингулярный поток представляет собою, очевидно, суперпозицию конечного или счетного числа элементарных потоков только что описанного типа, независимых между собою. Таким образом, для производящей функции общего пуассоновского сингулярного потока мы находим выражение

$$\Gamma(x, \alpha, \beta) = \exp\left\{(x-1) \sum_{\alpha \leq t_i < \beta} \alpha_i\right\}.$$

При этом необходимо допустить, что ряд

$$\sum_{0 \leq t_i < t} \alpha_i = C(t)$$

сходится при любом $t \geq 0$; мы имеем

$$\Gamma(x, \alpha, \beta) = \exp\{[C(\beta) - C(\alpha)](x-1)\}.$$

Составим теперь суперпозицию данного пуассоновского сингулярного потока с регулярным ординарным потоком (имеющим, согласно § 2, форму (3)), ведущую функцию которого мы обозначим через $\Lambda(t)$, а производящую функцию через

$$\Phi(x, \alpha, \beta) = \exp\{[\Lambda(\beta) - \Lambda(\alpha)](x-1)\}.$$

Эта суперпозиция будет, очевидно, иметь тот же тип производящей функции, что и ее компоненты, с $C(t) + \Lambda(t)$ в качестве ведущей функции; образуемый ею поток имеет, следовательно, форму (3). Мы видим, таким образом, что *суперпозиция двух взаимно независимых потоков, один из которых — регулярный ординарный, а другой — пуассоновский сингулярный с рядом значений параметра, сходящимся в любом конечном интервале, всегда является потоком вида (3).*

Рассмотрим теперь произвольный поток без последствий вида (3), где $\Lambda(t)$ означает некоторую неубывающую функцию [$\Lambda(t) = 0$ при $t < 0$]. Пусть $t_1, t_2, \dots$ занумерованные в любом порядке точки разрыва функции $\Lambda(t)$. Положим

$$\Lambda(t_i + 0) - \Lambda(t_i - 0) = \alpha_i \quad (i = 1, 2, \dots), \quad \sum_{0 \leq t_i < t} \alpha_i = C(t),$$

так что $\Lambda(t) - C(t) = \Lambda^*(t)$ непрерывная неубывающая функция. Положим далее

$$\begin{aligned}\exp\{[C(\beta) - C(\alpha)](x-1)\} &= \Gamma(x, \alpha, \beta), \\ \exp\{[\Lambda^*(\beta) - \Lambda^*(\alpha)](x-1)\} &= \Phi^*(x, \alpha, \beta), \\ \exp\{[\Lambda(\beta) - \Lambda(\alpha)](x-1)\} &= \Phi(x, \alpha, \beta),\end{aligned}$$

откуда

$$\Phi(x, \alpha, \beta) = \Gamma(x, \alpha, \beta) \Phi^*(x, \alpha, \beta).$$

Оба множителя правой части этого равенства представляют собою производящие функции потоков типа (3), ведущие функции которых соответственно равны $C(t)$ и $\Lambda^*(t)$. Так как функция $\Lambda^*(t)$ непрерывна, то второй из этих потоков — регулярный; а так как это поток типа (3), то в силу результатов § 2 он должен быть ординарным.

Рассмотрим теперь первый поток с производящей функцией $\Gamma(x, \alpha, \beta)$. Так как

$$C(\beta) - C(\alpha) = \sum_{\alpha \leq t_i < \beta} \alpha_i,$$

то

$$\Gamma(x, \alpha, \beta) = \prod_{\alpha \leq t_i < \beta} \exp\{\alpha_i(x-1)\};$$

отдельный множитель произведения; стоящего в правой части этого равенства, имеет вид

$$\exp\{\alpha_i(x-1)\} = e^{-\alpha_i} \sum_{k=0}^{\infty} \frac{\alpha_i^k}{k!} x^k,$$

и, следовательно, представляет собою производящую функцию распределения Пуассона с параметром α_i ; следовательно, $\Gamma(x, \alpha, \beta)$ есть производящая функция пуассоновского сингулярного потока со степенями t_i и параметрами $\alpha_i = \Lambda(t_i + 0) - \Lambda(t_i - 0)$, так что ряд

$$\sum_{0 \leq t_i < t} \alpha_i \leq \Lambda(t)$$

сходится при любом $t \geq 0$.

Мы приходим, таким образом, к заключению, что *любой поток типа (3) может быть представлен в виде суперпозиции двух взаимно независимых потоков, из которых один регулярный и ординарный, а другой — пуассоновский сингулярный со сходящимися в любом конечном интервале рядом параметров.*

Сопоставляя это с результатом, полученным выше, мы приходим к следующему предложению, которое и решает поставленную в настоящей статье задачу.

Теорема. Для того чтобы поток без последствия имел форму (3), необходимо и достаточно, чтобы он представлял собою суперпозицию двух взаимно независимых потоков, из которых один регулярный и ординарный, а другой — пуассоновский сингулярный, с рядом параметров, сходящимся в каждом конечном интервале.

Разложение, о котором здесь идет речь, непосредственно определяется ведущей функцией $\Lambda(t)$ данного потока: ступенями сингулярной компоненты служат точки разрыва функции $\Lambda(t)$, в качестве параметра α_i для ступени t_i фигурирует величина $\Lambda(t_i + 0) - \Lambda(t_i - 0)$; наконец, ведущими функциями компонент служат $C(t) = \sum_{t_i < t} \alpha_i$ (для сингулярной компоненты) и $\Lambda^*(t) = \Lambda(t) - C(t)$ (для регулярной компоненты).

ЛИТЕРАТУРА

- [1] А. Я. Хинчин, Математические методы теории массового обслуживания. Труды Матем. ин-та. АН СССР, 49 (1955), стр. 1—122.
- [2] А. Я. Хинчин, Теория вероятностей и ее применения, 1. вып. 1 (1956), стр. 3—17.

О ФОРМУЛАХ ЭРЛАНГА В ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ *)

1. Постановка задачи. Все изложенное в настоящей статье относится в равной мере к любым установкам массового использования, и лишь ради краткости мы будем пользоваться терминологией, принятой в приложениях теории вероятностей к телефону.

В теории массового обслуживания принято называть законом распределения $f(x)$ случайной величины ξ вероятность неравенства $\xi > x$; в этом смысле законы распределения и будут пониматься во всем дальнейшем.

Пусть мы имеем дело с телефонной установкой, имеющей n обслуживающих устройств, которые мы для краткости будем называть линиями. На эти линии поступают требования (вызовы). Если в момент такого вызова имеется свободная (незанятая) линия, то вызов занимает ее (одну из них, если свободных линий несколько). Период занятия линии одним вызовом называется разговором; по окончании

*) Теория вероятностей и ее применения, т. 7, вып. 3, 1962, стр. 330—335.

Рукопись настоящей статьи А. Я. Хинчина обнаружена мной в его бумагах 12.6.60 г. Рукопись была полностью подготовлена к печати: перепечатана на машинке, формулы вставлены рукой автора. Бумага от времени пожелтела. Судя по тому, что известная монография Хинчина «Математические методы теории массового обслуживания» была закончена летом 1954 г. и напечатана в первой половине 1955 г., а начиная с мая 1956 г. А. Я. Хинчин уже тяжело заболел, эта рукопись относится к периоду август 1954 г.—апрель 1956 г.

Основной результат настоящей работы был получен Б. А. Севастьяновым в работе «Эргодическая теорема для марковских процессов и ее приложение к телефонным системам с отказами» (Теория вероятностей и ее применения, т. 2, вып. 1, 1957, стр. 106—116).

Однако метод А. Я. Хинчина может оказаться полезным при дальнейших исследованиях. — Б. Г.

разговора линия освобождается и может быть занята новыми вызовами. Если в момент поступления вызова все n линий заняты, то вызов получает отказ (потерю) и все дальнейшее происходит так, как если бы этого вызова не было. Если в данный момент занято k из общего числа n линий ($k=0, 1, \dots, n$), то мы для краткости будем говорить, что система находится в состоянии k .

Одним из важнейших показателей качества обслуживания для данной установки служат вероятности различных ее состояний. Под вероятностью состояния k при этом всегда понимается доля времени, в течение которого система находится в этом состоянии. При этом имеется в виду, что промежуток времени T , в течение которого ведется наблюдение, очень велик. В частности, вероятность состояния n , т. е. доля времени, в течение которого поступающие вызовы получают отказы, называется вероятностью потери (или опасным временем). Очевидно, что вероятности состояний зависят как от природы поступающего потока вызовов, так и от закона распределения длительности разговоров. Поступающий поток вызовов обычно предполагается простейшим; это значит, что для любого момента времени a вероятность отсутствия вызовов в промежутке $(a, a+t)$ равна $e^{-\lambda t}$ (где λ — постоянное положительное число) и не зависит от всего предшествующего течения потока. Что касается длительности разговоров, то, прежде всего, предполагается, что длины различных разговоров не зависят ни друг от друга, ни от того, как протекает поток вызовов.

Пусть $F(t)$ — закон распределения длительности разговоров, т. е. доля тех разговоров, длительность которых превосходит t . Эрланг выводил свои известные формулы вероятностей состояний в предположении $F(t) = e^{-\beta t}$ ($\beta > 0$ — постоянная); это допущение, как известно, вообще значительно облегчает исследование вопросов теории массового обслуживания. Однако ввиду важности задачи позднее был сделан ряд попыток показать, что формулы, найденные Эрлангом, сохраняют силу при любом законе $F(t)$ (иногда этот закон подчиняется некоторым требованиям общего характера, например условию непрерывности)*). Во всех этих попытках

*) См. [1], [2].

используется сложный аналитический аппарат и, насколько мы можем судить, так и не приводит к окончательному решению задачи. Лишь в 1953 г. появилась работа Лундквиста [3], знаменующая собой некоторый сдвиг в этом направлении; с помощью нового, простого и элементарного метода Лундквисту удалось показать, что формулы Эрланга сохраняют силу и в случае, когда все разговоры имеют одну и ту же длину τ , т. е. в случае

$$F(t) = \begin{cases} 1 & (t < \tau), \\ 0 & (t \geq \tau). \end{cases}$$

Настоящая работа ставит себе целью показать, что некоторое усовершенствование метода Лундквиста позволяет установить формулы Эрланга и для любого закона распределения $F(t)$. При этом все рассуждение не только не усложняется, но, напротив, становится более кратким и обозримым.

2. Обозначения. Значение символов λ и $F(t)$ определено в п°1; обозначим через s среднюю длительность разговора, так что

$$s = - \int_0^{\infty} t dF(t) = \int_0^{\infty} F(t) dt.$$

Отрезок времени, в течение которого система находится в состоянии k , мы будем обозначать через Δ_k . Число всех отрезков Δ_k за время T обозначается через N_k . Очевидно, что каждый отрезок Δ_k начинается и кончается либо вызовом, либо освобождением линии. Обозначим через $A_k(B_k)$ долю отрезков Δ_k , начинающихся вызовом (освобождением); через $a_k(b_k)$ — долю отрезка Δ_k , кончающихся вызовом (освобождением), и, наконец, через Aa_k — долю отрезков Δ_k , начинающихся и кончающихся вызовом.

Ab_k , Ba_k , Bb_k определяются по аналогии очевидным образом.

Условимся говорить, что данный отрезок Δ_k *принадлежит типу A*, если он начинается с вызова. Тогда непосредственно ясно, что означает принадлежность отрезка Δ_k типам B , a , b , Aa , Ab , Ba , Bb .

Далее, обозначим через $M_k(MA_k, MB_k)$ среднюю длительность отрезка Δ_k (отрезка Δ_k типа A , отрезка Δ_k типа B), так что

$$M_k = A_k MA_k + B_k MB_k. \quad (1)$$

Наконец, пусть $[k]$ ($k=0, 1, \dots, n$) означает вероятность состояния k , т. е. отношение суммарной длительности всех отрезков Δ_k за (большой) промежуток времени T к длине T этого промежутка.

3. Элементарная статистика отрезков Δ_k . Из определений п°2 с непосредственной очевидностью вытекают соотношения ($0 \leq k \leq n$)

$$A_k + B_k = a_k + b_k = 1 \quad (2)$$

$$\left. \begin{aligned} Aa_k + Ab_k &= A_k, & Ba_k + Bb_k &= B_k, \\ Aa_k + Ba_k &= a_k, & Ab_k + Bb_k &= b_k. \end{aligned} \right\} \quad (3)$$

Далее, непосредственно очевидно, что между двумя последовательными отрезками Δ_k типа Aa должен обязательно встретиться отрезок Δ_k типа Bb , и обратно; из этого следует важное соотношение

$$Aa_k = Bb_k; \quad (4)$$

отсюда и из формул (3) следует

$$a_k = Aa_k + Ba_k = Bb_k + Ba_k = B_k \quad (5)$$

и аналогично

$$b_k = A_k.$$

4. Исходная рекуррентная формула. За каждым отрезком Δ_{k-1} ($1 \leq k \leq n$) типа a следует отрезок Δ_k типа A ; обратно, каждому отрезку Δ_k типа A предшествует некоторый отрезок Δ_{k-1} типа a . Отсюда

$$N_{k-1}a_{k-1} = N_k A_k,$$

или в силу (5)

$$N_{k-1}B_{k-1} = N_k A_k. \quad (6)$$

Так как суммарная длительность всех отрезков Δ_k равна $T[k]$, то

$$M_k = \frac{T[k]}{N_k}, \quad N_k = \frac{T[k]}{M_k} \quad (k=0, 1, \dots, n).$$

Поэтому из (6) следует

$$\frac{B_{k-1}T[k-1]}{M_{k-1}} = \frac{A_k T[k]}{M_k},$$

или

$$\frac{A_k}{M_k}[k] = \frac{B_{k-1}}{M_{k-1}}[k-1] \quad (k=1, 2, \dots, n). \quad (7)$$

Эта элементарная рекуррентная формула и служит отправным пунктом при выводе формул Эрланга. Для ее применения необходимо найти величины A_k , B_k и M_k . К их постепенному отысканию мы теперь и переходим.

5. Закон распределения оставшейся части разговора. Пусть в некоторый произвольно выбранный момент времени мы застали ведущийся разговор (наряду, быть может, с другими разговорами). Обозначим через $\varphi(t)$ вероятность того, что по истечении t сек этот разговор еще не будет закончен. Чтобы подойти к определению функций $\varphi(t)$, разобьем отрезок времени $(0, T)$ на очень малые части одинаковой длины τ (ячейки). Отрезок какого-либо разговора, приходящийся на одну такую ячейку длины τ , условимся называть *элементом*. Если под нормальным элементом понимать такой, у которого содержащий его разговор по истечении t сек еще не будет закончен, то $\varphi(t)$ можно статистически определить как отношение числа нормальных элементов к числу всех элементов (точнее, как предел этого отношения при $\tau \rightarrow 0$). Число всех элементов равно суммарной длительности всех ведущихся в отрезке $(0, T)$ разговоров, разделенной на τ , т. е. равно

$$\frac{Ns}{\tau}, \quad (8)$$

где N — число всех разговоров за время T . Подсчитаем теперь, сколько среди этих элементов будет нормальных.

Число разговоров в отрезке $(0, T)$, длина которых заключена между u и $u + du$, будет

$$N[F(u) - F(u + du)] = -NdF(u).$$

При $u < t$ разговор такой длины, очевидно, нормальных элементов содержать не будет. Если же $u > t$, то нормальными элементами будут те, которые отстоят менее чем на $u - t$ от начала разговора (более чем на t от его конца). Число нормальных элементов для одного разговора длины $u > t$ поэтому равно $(u - t)/\tau$, а число таких элементов для

всех разговоров длины u равно

$$-\frac{N(u-t)dF(u)}{\tau} \quad (u > t).$$

Суммируя по u это выражение, мы находим, что число всех нормальных элементов равно

$$-\frac{N}{\tau} \int_t^{\infty} (u-t) dF(u); \quad (9)$$

для функции $\varphi(t)$, являющейся отношением величин (9) и (8), мы отсюда находим выражение

$$\varphi(t) = -\frac{1}{s} \int_t^{\infty} (u-t) dF(u).$$

Интеграция по частям дает

$$\varphi(t) = \frac{1}{s} \int_t^{\infty} F(u) du. \quad (10)$$

6. Законы распределения длин отрезков Δ_k различных типов. Рассмотрим какой-либо отрезок Δ_k типа A и допустим, что $0 < k < n$ (отрезков Δ_0 типа A не существует, а случай $k = n$ требует особого рассмотрения, которое будет проведено в дальнейшем). Для того чтобы этот отрезок имел длину $> t$, необходимо и достаточно выполнение следующих трех условий:

1. В течение t сек после начала Δ_k не поступает ни одного вызова; вероятность этого условия равна $e^{-\lambda t}$.

2. Разговор того вызова, которым начинается Δ_k , имеет длительность $> t$; вероятность этого условия равна $F(t)$.

3. Каждый из тех $k-1$ разговоров, которые уже велись в начальный момент Δ_k , продолжается еще $> t$ сек. Вероятность этого условия ввиду взаимной независимости длительностей различных разговоров равна $[\varphi(t)]^{k-1}$.

Так как эти три условия взаимно независимы, то для отрезка Δ_k типа A вероятность длины $> t$ (закон распределения) равна

$$\psi_A(t) = e^{-\lambda t} F(t) [\varphi(t)]^{k-1} \quad (0 < k < n).$$

Для отрезка Δ_k типа B дело обстоит несколько иначе, так как здесь для всех k ведущихся разговоров время их начала неизвестно; поэтому рассуждение, аналогичное только что нами проведенному, для длин отрезков Δ_k типа B дает, как легко видеть, закон распределения

$$\psi_B(t) = e^{-\lambda t} [\varphi(t)]^k \quad (0 \leq k < n)$$

(эта формула пригодна и при $k=0$).

С другой стороны, нас будет интересовать вероятность того, что отрезок Δ_k того или другого типа закончится поступлением нового вызова в возрасте между t и $t+dt$. Чтобы это произошло, необходимо выполнение следующих двух условий: 1) ни один из k разговоров, ведущихся в данном Δ_k , не должен закончиться за время t после начала Δ_k ; вероятность этого условия равна, как мы уже видели, $F(t) [\varphi(t)]^{k-1}$ для Δ_k типа A ($0 < k < n$) и $[\varphi(t)]^k$ для Δ_k типа B ($0 \leq k < n$); 2) первый после начала данного Δ_k вызов должен поступить через промежуток времени, заключенный между t и $t+dt$; вероятность этого условия равна $\lambda e^{-\lambda t} dt$. Таким образом, вероятность закончиться вызовом в возрасте между t и $t+dt$ равна

$$\begin{aligned} \lambda e^{-\lambda t} F(t) [\varphi(t)]^{k-1} dt &= \lambda \psi_A(t) dt \quad (0 < k < n) \text{ (для } \Delta_k \text{ типа } A), \\ \lambda e^{-\lambda t} [\varphi(t)]^k dt &= \lambda \psi_B(t) dt \quad (0 \leq k < n) \text{ (для } \Delta_k \text{ типа } B). \end{aligned}$$

Вероятность же закончиться вызовом независимо от возраста (т. е. принадлежать типу a) для отрезка Δ_k того или другого типа A или B получится интегрированием того или другого из только что написанных выражений по t от 0 до ∞ . Но эта вероятность есть для типа A условная вероятность принадлежать типу Aa , а для типа B — условная вероятность принадлежать типу Ba . Эти же две условные вероятности в принятых нами обозначениях выражаются соответственно в виде Aa_k/A_k и Ba_k/B_k . Таким образом,

$$\left. \begin{aligned} \frac{Aa_k}{A_k} &= \lambda \int_0^{\infty} \psi_A(t) dt \quad (0 < k < n), \\ \frac{Ba_k}{B_k} &= \lambda \int_0^{\infty} \psi_B(t) dt \quad (0 \leq k < n). \end{aligned} \right\} \quad (11)$$

7. Определение чисел A_k , B_k и M_k . Мы можем теперь перейти к определению A_k , B_k и M_k , входящих в рекуррентное соотношение (7). При этом мы по-прежнему предполагаем $k < n$.

Так как $\psi_A(t)$ есть закон распределения длин отрезков Δ_k типа A, то среднее значение MA_k этих длин равно

$$MA_k = - \int_0^{\infty} t d\psi_A(t) = \int_0^{\infty} \psi_A(t) dt \quad (0 < k < n),$$

и аналогично

$$MB_k = \int_0^{\infty} \psi_B(t) dt \quad (0 \leq k < n);$$

поэтому соотношения (11) дают

$$\frac{Aa_k}{A_k} = \lambda MA_k \quad (0 < k < n), \quad \frac{Ba_k}{B_k} = \lambda MB_k \quad (0 \leq k < n). \quad (12)$$

Из этих же формул в силу (4) получаем

$$\lambda MB_k = \frac{Ba_k}{B_k} = 1 - \frac{Bb_k}{B_k} = 1 - \frac{Aa_k}{B_k} = 1 - \lambda \frac{A_k}{B_k} MA_k \quad (0 \leq k < n). \quad (13)$$

С другой стороны, интегрируя по частям и формула (10) дают

$$\begin{aligned} \lambda MB_k &= \lambda \int_0^{\infty} \psi_B(t) dt = \int_0^{\infty} \lambda e^{-\lambda t} [\varphi(t)]^k dt = \\ &= 1 + k \int_0^{\infty} e^{-\lambda t} [\varphi(t)]^{k-1} d\varphi(t) = \\ &= 1 - \frac{k}{s} \int_0^{\infty} e^{-\lambda t} [\varphi(t)]^{k-1} F(t) dt = \\ &= 1 - \frac{k}{s} \int_0^{\infty} \psi_A(t) dt = 1 - \frac{k}{s} MA_k \quad (0 \leq k < n). \quad (14) \end{aligned}$$

Приравнявая друг другу правые части равенств (13) и (14), мы находим

$$\frac{A_k}{B_k} = \frac{k}{s\lambda} \quad (0 \leq k < n);$$

это же в силу (2) дает

$$A_k = \frac{k}{k + s\lambda}, \quad B_k = \frac{s\lambda}{k + s\lambda} \quad (0 \leq k < n).$$

Наконец, в силу (12) и (5) формула (1) дает при $0 \leq k < n$
 $M_k = A_k M A_k + B_k M B_k = \frac{1}{\lambda} [A a_k + B a_k] = \frac{a_k}{\lambda} = \frac{B_k}{\lambda} = \frac{s}{k + s\lambda}$;
 мы нашли, таким образом, числа A_k , B_k и M_k для всех $k < n$.

Нам остается рассмотреть случай $k = n$. Легко видеть, что в этом случае

$$A_n = 1, \quad B_n = 0, \quad \psi_A(t) = F(t) [\varphi(t)]^{n-1},$$

вследствие чего в силу (10)

$$\begin{aligned} M_n = M A_n &= \int_0^{\infty} \psi_A(t) dt = \int_0^{\infty} [\varphi(t)]^{n-1} F(t) dt = \\ &= -s \int_0^{\infty} [\varphi(t)]^{n-1} d\varphi(t) = \frac{s}{n}. \end{aligned}$$

8. Формулы Эрланга. Подставляя в рекуррентное соотношение (7) вместо A_k , B_{k-1} , M_k и M_{k-1} найденные нами значения этих величин, мы легко получаем (случай $k = n$ не составляет исключения)

$$[k] = \frac{\lambda s}{k} [k-1] \quad (1 \leq k \leq n),$$

откуда

$$[k] = \frac{(\lambda s)^k}{k!} [0] \quad (0 \leq k \leq n).$$

Присоединяя же сюда нормирующее соотношение

$$\sum_{k=0}^n [k] = 1,$$

находим

$$[0] = \left\{ \sum_{k=0}^n \frac{(\lambda s)^k}{k!} \right\}^{-1},$$

что и дает известные формулы Эрланга.

Впрочем, ничто в предыдущем не мешает нам считать число линий n бесконечно большим. В этом случае мы получаем также хорошо известные формулы Пуассона

$$[k] = e^{-\lambda s} \frac{(\lambda s)^k}{k!} \quad (k = 0, 1, 2, \dots).$$

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

- [1] A. E. Vaulot, Rev. gén. de l'électricité, 1927, 22, № 26, стр. 1164—1171.
 - [2] C. Palm, Ericsson Technics, 1938, 6, № 4, стр. 39—58.
 - [3] K. Lundkvist, Ericsson Technics, 1953, 9, № 2, стр. 111—140.
-

§ 1. Постановка задачи и предварительные замечания

В настоящей работе исследуется вопрос о числе потерь и времени ожидания в случае двух абонентов, пользующихся одним и тем же проводом. Для простоты предполагаем, что все разговоры имеют одинаковую длительность, которую мы будем обозначать через T (единица времени произвольна, мы будем условно называть ее часом). Это допущение может быть, конечно, оправдано лишь в самом грубом приближении, так как, по-видимому, дисперсия длительности разговоров может оказывать довольно существенное влияние на интересующие нас величины; однако развиваемый нами метод позволяет принципиально решить все поставленные вопросы и в случае любого закона распределения длительности разговоров; различие заключается лишь в том, что в случае переменной длительности расчеты становятся значительно сложнее; только по этой причине мы принимаем в настоящем исследовании (основная цель которого — выработка метода) длительность всех разговоров одинаковой. Далее, пусть каждый из двух абонентов в течение часа производит n_1 вызовов и сам получает n_2 вызовов (мы допускаем, таким образом, что эти числа для обоих абонентов одинаковы; если бы они были различны, это вызвало бы в дальнейших расчетах лишь самые незначительные усложнения). Величинами n_1 , n_2 и T исчерпываются данные задачи: все остальные числа, с которыми мы будем встречаться, должны быть через них выражены.

*) Рукопись настоящей статьи А. Я. Хиичина была обнаружена мною в бумагах автора. Она написана в 1934 г., но не была опубликована. В настоящее время эта тематика приобретает актуальный интерес.— Б. Г.

Мы допускаем, что в случае, когда вызов поступает в момент занятости провода, вызывающий (абонент или посторонний) ожидает его освобождения, после чего приступает к разговору; мы допускаем далее, что в случае скопления нескольких вызовов они обслуживаются проводом в порядке очереди; это последнее допущение, по-видимому, не оказывает никакого влияния на число потерь и среднее время ожидания и потому может быть сделано без всяких оговорок.

Если в момент t один из абонентов свободен, то вероятность того, что он произведет вызов до наступления момента $t + dt$, с точностью до малых высших порядков равна λdt , где λ — постоянная, которая в дальнейшем должна быть определена из данных нашей задачи. Отсюда, как известно, следует, что если абонент в данный момент свободен, то вероятность того, что он не произведет вызова в течение конечного промежутка времени t , равна $e^{-\lambda t}$.

Что касается вызовов, возникающих извне, то мы, как обычно, предполагаем их режим совершенно независимым от имеющегося в данный момент состояния скученности. Как известно, это имеет своим следствием так называемый закон Пуассона: вероятность появления j вызовов в течение промежутка времени t выражается формулой

$$e^{-\nu} \frac{(\nu t)^j}{j!}.$$

Здесь ν означает (среднее) число внешних вызовов в единицу времени и, следовательно, $\nu = 2n_2$.

Нам важно иметь удобное обозначение для состояния скученности в данный момент; условимся обозначать через a ожидающего абонента и через e — ожидающий внешний вызов; запись $eaee$ будет означать, что в данный момент имеется четверо ожидающих, причем первым в очереди стоит внешний вызов, вторым — абонент, а затем — еще два внешних вызова.

Нам надо также иметь обозначение для вероятности того, что в течение данного разговора, в начале которого скученность известна, произойдут те или другие определенные вызовы; так как на число этих вызовов может оказать влияние только наличие или отсутствие в начале разговора ожидающего абонента и так как более одного ожидающего абонента, очевидно, быть не может, то все разговоры мы

должны разделить на два класса: 1) те, в начале которых нет ожидающего абонента, и 2) те, в начале которых имеется один ожидающий абонент; эти два класса мы будем различать соответственно верхними индексами 0 и 1; мы введем следующие обозначения:

v_0^0 — вероятность того, что в течение разговора первого класса не поступит ни одного вызова;

v_a^0 — вероятность того, что в течение разговора первого класса поступит один вызов абонента (и ни одного внешнего вызова);

v_e^0 — вероятность того, что в течение разговора первого класса поступит один вызов внешний;

v_{ae}^0 — вероятность того, что в течение разговора первого класса поступит сначала вызов абонента, а затем — внешний;

v_{ea}^0 — вероятность того, что в течение разговоров первого класса поступят те же вызовы в обратном порядке, и т. д.;

v_0^1 — вероятность того, что в течение разговора второго класса не поступит ни одного вызова;

v_e^1 — вероятность того, что в течение разговора второго класса поступит один внешний вызов;

v_{ee}^1 — вероятность того, что в течение разговора второго класса поступят два внешних вызова, и т. д.

(Очевидно, что в течение разговора второго класса абонент вызывать не может). Все эти величины легко выражаются через данные нашей задачи и через величину κ (которую, как уже замечено, мы впоследствии также выразим через эти данные). В самом деле, полагая для краткости

$$e^{-\alpha T} = \alpha,$$

$$e^{-\nu T} \frac{(\nu T)'}{j!} = \tau j,$$

мы, очевидно, будем иметь:

$$v_0^0 = \alpha \tau_0,$$

$$v_e^1 = \tau_1;$$

$$v_a^0 = (1 - \alpha) \tau_0,$$

$$v_{ee}^1 = \tau_2,$$

$$v_e^0 = \alpha \tau_1,$$

$$v_{eec}^1 = \tau_3 \text{ и т. д.}$$

$$v_0^1 = \tau_0,$$

Несколько сложнее выражаются такие вероятности, как v_{ae}^0 и v_{ea}^0 ; однако и этот расчет не вызывает существенных затруднений. Вероятность того, что в течение данного разговора произойдет один вызов, и притом как раз в элементарный промежуток времени между t и $t + dt$ (считая от начала разговора), равна, очевидно, $\tau_1 \frac{dt}{T}$; вероятность того, что абонент, свободный в начале разговора, не произведет вызова до момента t , но произведет его в промежутке между t и T , равна

$$e^{-\alpha t} [1 - e^{-\alpha (T-t)}],$$

так что

$$\begin{aligned} v_{ea}^0 &= \frac{\tau_1}{T} \int_0^T e^{-\alpha t} [1 - e^{-\alpha (T-t)}] dt = \\ &= \frac{\tau_1 (1 - \alpha)}{\alpha T} - \alpha \tau_1; \end{aligned}$$

аналогично легко находим:

$$v_{ae}^0 = \tau_1 - \frac{\tau_1}{\alpha T} (1 - \alpha).$$

Подобным образом могут быть вычислены и более сложные вероятности того же типа.

Наконец, мы переходим к обозначению величин, являющихся основным предметом нашего исследования. Обозначим через π_0 относительное число разговоров, начинающихся без ожидания, или — что то же — относительное число вызовов, застающих провод свободным. Относительное число потерь (точнее — задержанных вызовов) равно тогда $1 - \pi_0$; таким образом, определение величины π_0 , естественно, является основной задачей нашего исследования.

Обозначим далее через π_a относительное число таких разговоров, перед началом которых ожидающим является один абонент; через π_e — относительное число таких разговоров, перед началом которых ожидающим является один внешний вызов; в обоих случаях непосредственно после начала данного разговора ожидающих не имеется. Вообще, через $\pi_{aee} \dots$ мы будем обозначать относительное число таких разговоров, непосредственно перед началом которых скученность

характеризуется схемой $aee\dots$; в момент начала такого разговора первый охлаждающий выбывает из очереди, и следовательно, непосредственно после начала разговора скученность определяется схемой $ee\dots$

§ 2. Основная система уравнений

Как уже сказано, нашей основной целью является определение величины π_0 . Однако, как и во всех подобного рода задачах, эту величину удастся определить лишь из системы линейных уравнений, в которую наряду с π_0 в качестве неизвестных входят все величины: $\pi_a, \pi_e, \pi_{ae}, \pi_{ea}, \pi_{aee}, \dots$, т. е. из системы бесконечного числа линейных уравнений с бесконечным числом неизвестных. К составлению этой системы мы теперь и переходим.

Для того чтобы некоторый разговор начался без предварительного ожидания (вероятность π_0), необходимо и достаточно, чтобы 1) в начале предыдущего разговора не было ни одного ожидающего (вероятность $\pi_0 + \pi_a + \pi_e$) и 2) в течение этого предыдущего разговора не поступило ни одного вызова (вероятность v_0^0); поэтому, полагая для краткости

$$\pi_0 + \pi_a + \pi_e = x,$$

мы будем иметь:

$$\pi_0 = v_0^0 x;$$

это — первое уравнение нашей системы.

Далее, для того чтобы некоторый разговор начался в момент, когда скученность характеризуется схемой a (вероятность π_a), необходимо и достаточно выполнение одной из следующих двух предпосылок — либо I: 1) предшествующий разговор вначале протекает без ожидающих (вероятность x) и 2) в течение этого разговора происходит один вызов от абонента (вероятность v_a^0), либо II: 1) предшествующий разговор вначале протекает при одном ожидающем абоненте (вероятность π_{ea}) и 2) в течение этого разговора никаких вызовов не происходит; отсюда

$$\pi_a = v_a^0 x + v_0^1 \pi_{ea}$$

Это — второе уравнение нашей системы.

Совершенно аналогичными рассуждениями мы получаем и дальнейшие уравнения нашей системы; вот несколько из них:

$$\begin{aligned}\pi_0 &= v_0^0 x + v_0^0 \pi_{ee} + v_0^0 \pi_{ae}, \\ \pi_{ea} &= v_{ea}^0 x + v_a^0 \pi_{ee} + v_a^0 \pi_{ae} + v_0^0 \pi_{eee}, \\ \pi_{ae} &= v_{ae}^0 x + v_e^0 \pi_{ea} + v_0^0 \pi_{eee}, \\ \pi_{ee} &= v_{ee}^0 x + v_e^0 \pi_{ae} + v_e^0 \pi_{ea} + v_0^0 (\pi_{eee} + \pi_{eee})\end{aligned}$$

и т. д.

Прежде всего заметим, что искомая величина π_0 во все уравнения, кроме первого, входит только в комбинации $\pi_0 + \pi_a + \pi_e = x$; поэтому целесообразно ввести x в качестве новой неизвестной вместо π_0 ; при этом мы сначала выбросим первое уравнение нашей системы, поставив своей основной целью определение x ; когда это будет сделано, мы вернемся к этому первому уравнению, чтобы определить из него π_0 .

Далее, заметим, что наша система, будучи однородной, позволит нам в лучшем случае определить взаимные отношения чисел π ; чтобы найти сами эти числа, мы должны, естественно, воспользоваться нормирующим соотношением

$$\pi_0 + \pi_a + \pi_e + \pi_{ae} + \pi_{ea} + \dots = 1.$$

Таким образом, наша система получает следующий вид:

$$x + \pi_{ae} + \pi_{ea} + \pi_{ee} + \dots = 1,$$

$$\left. \begin{aligned}\pi_a &= v_a^0 x + v_0^0 \pi_{ea}, \\ \pi_e &= v_e^0 x + v_0^0 \pi_{ee} + v_0^0 \pi_{ae}, \\ \pi_{ea} &= v_{ea}^0 x + v_a^0 \pi_{ee} + v_a^0 \pi_{ae} + v_0^0 \pi_{eee}, \\ \pi_{ae} &= v_{ae}^0 x + v_e^0 \pi_{ea} + v_0^0 \pi_{eee}, \\ \pi_{ee} &= v_{ee}^0 x + v_e^0 \pi_{ae} + v_e^0 \pi_{ea} + v_0^0 (\pi_{eee} + \pi_{eee}).\end{aligned} \right\} \quad (1)$$

Эта система, в отличие от других, встречаемых в более простых задачах телефонии, не может быть решена методом последовательного (рекуррентного) определения неизвестных;

точное ее решение представляет поэтому значительные трудности. В дальнейшем мы увидим, что для практических целей достаточно дать приближенное решение, получаемое весьма просто.

К системе (1) надо присоединить еще одно добавочное уравнение

$$\pi_0 = v_0^0 x, \quad (2)$$

определяющее величину π_0 .

§ 3. Приближенное решение основной системы

Все величины, входящие в нашу основную систему (1) (как данные, так и искомые), можно рассматривать как функции от T ; на практике T обычно представляет собой небольшую дробь (примерно $1/_{10}$). При $T \rightarrow 0$ величины v_0^0 и v_0^1 стремятся к единице, а величины v с буквенными нижними индексами — к нулю; при этом v_a и v_e являются (независимо от верхних индексов) малыми первого порядка, v_{ae} , v_{ea} , v_{ee} — второго и т. д. Аналогично этому $\pi_0 \rightarrow 1$, а все π с буквенными индексами бесконечно малы, причем порядок малости совпадает с числом нижних индексов.

Это обстоятельство, естественно, наводит нас на мысль искать приближенное решение системы (1), разлагая все данные и искомые величины по степеням T и ограничиваясь в этом разложении членами того или иного порядка. В дальнейшем мы проведем все вычисления, пренебрегая всеми, содержащими третью и более высокие степени T . Первым следствием этого соглашения является то, что система уравнений становится конечной, причем и в оставшихся уравнениях ряд членов исчезает. Первое, четвертое, пятое и шестое уравнения обращаются в систему:

$$\begin{aligned} x + \pi_{ae} + \pi_{ea} + \pi_{ee} &= 1, \\ \pi_{ea} &= v_{ea}^0 x, \\ \pi_{ae} &= v_{ae}^0 x, \\ \pi_{ee} &= v_{ee}^0 x, \end{aligned}$$

которая может быть решена без затруднений; мы получаем:

$$x = \frac{1}{1 + v_{ae}^0 + v_{ea}^0 + v_{ee}^0} = \frac{1}{1 + \tau_1(1 - \alpha) + \alpha\tau_2},$$

и, следовательно, с точностью до малых второго порядка

$$x = 1 - \tau_1(1 - \alpha) - \tau_2\alpha;$$

уравнение (2) дает нам поэтому

$$\pi_0 = \alpha\tau_0 [1 - \tau_1(1 - \alpha) - \tau_2\alpha].$$

Если теперь принять во внимание, что

$$\alpha = e^{-\kappa T} = 1 - \kappa T + \frac{\kappa^2 T^2}{2} + o(T^3),$$

$$\tau_0 = e^{-\nu T} = 1 - \nu T + \frac{\nu^2 T^2}{2} + o(T^3),$$

$$\tau_1 = \nu T e^{-\nu T} = \nu T - \frac{\nu^2 T^2}{2} + o(T^3),$$

$$\tau_2 = \frac{\nu^2 T^2}{2} e^{-\nu T} = \frac{\nu^2 T^2}{2} + o(T^3),$$

то мы после элементарных вычислений получаем с точностью до малых второго порядка

$$\pi_0 = 1 - (\nu + \kappa) T + \frac{1}{2} \kappa^2 T^2. \quad (3)$$

Однако эта формула еще не решает нашей задачи, как и вообще, с помощью одних только уравнений (1) и (2) ее невозможно решить. Дело в том, что если величина ν нам прямо дана ($\nu = 2n_2$), то величина κ , входящая в формулу (3), нам неизвестна; эта величина не принадлежит к числу априорных данных, а напротив, сама является функцией создающегося режима. Поэтому мы обращаемся к выводу второго соотношения, связывающего между собой величины π_0 и κ .

Если в некоторый момент t провод свободен, то вероятность появления вызова со стороны одного (определенного) абонента до момента $t + dt$ есть κdt ; вероятность же появления за этот промежуток времени внешнего вызова есть νdt ; таким образом, вероятность того, что к моменту $t + dt$ провод окажется занятым (при условии, что в момент t он был свободен), составляет $(2\kappa + \nu) dt$. С помощью обычного рассуждения мы отсюда находим: если в момент 0 провод был

свободен, то вероятность того, что ближайшее занятие его произойдет в промежутке времени между t и $t + dt$, равна

$$e^{-(2\kappa + \nu)t} (2\kappa + \nu) dt,$$

а это позволяет найти среднюю величину периода свободы:

$$\beta = \int_0^{\infty} t e^{-(2\kappa + \nu)t} (2\kappa + \nu) dt = \frac{1}{2\kappa + \nu};$$

но общее число периодов свободы, очевидно, равно числу разговоров, начинающихся без ожидания, т. е. $2(n_1 + n_2)\pi_0$. Поэтому суммарная длительность периодов свободы, если положить $n_1 + n_2 = n$, составит

$$\frac{2n\pi_0}{2\kappa + \nu},$$

а так как суммарная длительность всех разговоров есть $2nT$, то мы, очевидно, имеем:

$$\frac{2n\pi_0}{2\kappa + \nu} + 2nT = 1.$$

Это и дает нам искомое второе соотношение между π_0 и κ ; отсюда

$$\pi_0 = (2\kappa + \nu) \left(\frac{1}{2n} - T \right);$$

сопоставляя это с формулой (3), мы находим:

$$(2\kappa + \nu) \left(\frac{1}{2n} - T \right) = 1 - (\nu + \kappa) T + \frac{1}{2} \kappa^2 T^2,$$

а подстановка этого выражения в формулу (3) приводит нас к окончательному результату:

$$\pi_0 = 1 - (n_1 + 2n_2) T - (n_1 + 2n_2) \frac{n_1}{2} T^2. \quad (4)$$

Числовой пример: $T = 0,03$; $n_1 = 2$; $n_2 = 2,5$;

$$1 - \pi_0 = 0,2163 = 21,63\%.$$

§ 4. Вычисление среднего времени ожидания

Найти среднее время ожидания в данной схеме гораздо труднее, чем в обычных схемах телефонии. Мы ограничимся расчетом приближенной формулы, аналогичной той, которую мы нашли для числа потерь, и для практических целей вполне достаточной.

Обозначим через $P(k)$ относительное время, в течение которого имеется k ожидающих (абонентов или внешних вызовов — безразлично). Найдем сначала приближенное выражение для $P(0)$ — времени, протекающего в отсутствие ожидающих. Очевидно, что $P(0)$ складывается из следующих компонент:

1) время, когда провод свободен $\sim 1 - 2\pi T$;

2) время, когда провод занят, но ожидающих нет; это время мы должны вычислить.

Заметим, что относительное число разговоров, в начале которых не имеется ожидающих, равно

$$x = \pi_0 + \pi_a + \pi_e.$$

Если в момент 0 начался разговор этого типа, то вероятность того, что в течение малого промежутка dt последует вызов, равна $(\kappa + \nu) dt$, а потому, как известно, вероятность того, что до момента t вызова не последует, равна $e^{-(\kappa + \nu)t}$ ¹⁾; вероятность того, что ближайший вызов последует в промежутке $(t, t + dt)$, равна

$$e^{-(\kappa + \nu)t} (\kappa + \nu) dt.$$

Поэтому математическое ожидание той части разговора, которая протекает в отсутствие ожидающих, равно

$$\int_0^T e^{-(\kappa + \nu)t} (\kappa + \nu) t dt + T e^{-(\kappa + \nu)T},$$

что по вычислении дает

$$T - \frac{\nu + \kappa}{2} T^2 + o(T^3).$$

¹⁾ В эти формулы входят κ , а не 2κ (как прежде) потому, что один из абонентов занят разговором, и вызывать может только другой.

Таким образом, суммарная длительность тех частей разговоров, в течение которых не имеется ожидающих, равна с точностью до малых второго порядка

$$2nx \left\{ T - \frac{v+x}{2} T^2 \right\} = 2nTx \left\{ 1 - \frac{v+x}{2} T \right\};$$

принимая же во внимание сказанное выше, мы находим:

$$P(0) = 1 - 2nT + 2nTx \left\{ 1 - \frac{v+x}{2} T \right\}. \quad (5)$$

Теперь мы должны были бы перейти к вычислению величины $P(1)$; это можно произвести теми же методами, хотя и несравненно сложнее. Однако нетрудно сообразить и без всяких вычислений, как должно выразиться $P(1)$ с точностью, какую мы приняли. В самом деле, обозначая через $P(>1)$ время $\sum_{k=2}^{\infty} P(k)$, в течение которого число ожидающих превосходит единицу, мы непосредственно видим, что все это время принадлежит таким разговорам, при окончании которых (т. е. перед началом следующего разговора) число ожидающих не менее чем 2. Относительное число таких разговоров равно $1-x$ и есть, как мы знаем, малая второго порядка, а так как суммарная длительность таких разговоров получается умножением их числа на T , то она есть $n(1-x)T$; следовательно, она (а тем более и не превышающая ее величина $P(>1)$) есть малая третьего порядка. Но если так, то с точностью до малых второго порядка мы должны иметь:

$$P(0) + P(1) = 1,$$

или в силу формулы (5)

$$P(1) = 2n(1-x)T + n(v+x)T^2x,$$

или проще

$$P(1) + n(v+x)T^2x,$$

так как $2n(1-x)T$ есть малая третьего порядка. Но, с другой стороны, очевидно, что, обозначая через γ среднее время ожидания, мы должны иметь:

$$2n\gamma = P(1),$$

откуда

$$\gamma = \frac{v + x}{2} x T^2;$$

в пределах требуемой точности мы вправе положить $x = 1$ и $x = n_1$, что дает

$$\gamma = \left(\frac{n_1}{2} + n_2 \right) T^2.$$

В случае числового примера § 3 мы находим:

$$\gamma = 0,00315 \text{ часа} = 11,34 \text{ секунды.}$$

О НЕКОТОРЫХ ПОСТАНОВКАХ ЗАДАЧ И РЕЗУЛЬТАТАХ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

Б. В. Гнеденко

Современи выхода в свет первого издания монографии А. Я. Хинчина по теории массового обслуживания интерес к этому направлению исследований неизмеримо возрос как в Советском Союзе, так и за его пределами. За этот срок появились многочисленные оригинальные исследования, посвященные решению как чисто прикладных, так и теоретических вопросов. Методы исследования стали более разнообразны и гибки. Одновременно круг применений теории стал значительно более широким. Если в двадцатые годы отдельные вопросы теории массового обслуживания возникали лишь в практике организации телефонных сетей, то теперь к вопросам телефонного дела добавились разнообразнейшие и интересные задачи ядерной физики, автоматизации производства, эксплуатации аэропортов, морских причалов, автомобильных хозяйств, железнодорожных станций, больниц, ремонтных пунктов, предприятий массового обслуживания населения (магазинов, пунктов проката, билетных касс и пр.). В связи с таким увеличением прикладной роли теории массового обслуживания за последние несколько лет появились хорошие обзоры, приспособленные преимущественно к нуждам практиков. В первую очередь при этом, пожалуй, следует назвать обзорную статью Т. Саати [1], позднее перепечатанную в несколько измененном виде в его книге [2], посвященной изложению содержания новой научной дисциплины, известной под названием исследования операций. Более математический характер носят статья Л. Такача [1] и обзор, составивший главу 9 недавно появившейся книги Баруча Райда [1].

Из довольно многочисленных теперь монографий по теории массового обслуживания хотелось бы упомянуть книги Авондо

Бодино и Бромбилла [1], Кокса и В. Смита [1], Л. Такача [2], Р. Сиски [1], Т. Саати [3], Дж. Риордэна [1]. Каждая из названных книг представляет известный интерес, однако в первую очередь хотелось бы выделить среди них монографии Р. Сиски и Т. Саати. Первая из них представляет собой подробную сводку результатов по теории массового обслуживания, связанных с задачами телефонного дела. Вторая книга ставит перед собой более широкие цели — изложение основных результатов, полученных относительно важнейших в теоретическом и прикладном отношениях задач теории массового обслуживания.

В существующей монографической литературе по теории массового обслуживания книга А. Я. Хинчина занимает значительное место, поскольку в ней впервые было систематически изучено строение входящего потока требований, а также распределение времени ожидания для системы с очередью при обслуживании ее одним прибором. Заметим, что именно в этой монографии А. Я. Хинчин изложил свои достаточные условия близости суммарного потока, слагаемые которого независимы и равномерно малы, к простейшему потоку. По сути дела, как это выяснил ученик А. Я. Хинчина Г. А. Ососков, условия Хинчина являются и необходимыми.

Ниже мы изложим некоторые вопросы, решению которых последнее время уделялось значительное внимание.

С точки зрения математика теория массового обслуживания может быть включена в общую теорию случайных процессов и полей. При этом интересующие ее процессы будут весьма частного вида. Однако при таком подходе многие практически важные задачи могут приобретать весьма причудливый характер. В результате целесообразно выделить новый раздел теории случайных процессов, разрешив в нем собственную терминологию, способствующую естественной формулировке специфических проблем. Получается ситуация, близко напоминающая собой то, что произошло с самой теорией вероятностей: формально ее можно превратить в раздел теории функций действительного переменного, однако при этом будет потеряна специфика теории вероятностей, столь ценная как для самой математики, так и для многообразных применений. В теории массового обслуживания теория случайных процессов становится орудием исследования, не подмывая при этом под себя специфику возникающих вопросов.

Первичным понятием теории массового обслуживания, изучению которого большое внимание уделил А. Я. Хинчин, является понятие потока требований, поступающих на систему обслуживания. Кратко мы будем называть его входящим потоком. В том виде, в каком понятие потока до сих пор встречалось в литературе, оно укладывается в общее понятие случайного процесса. Это обстоятельство четко отмечено в § 6 книги Хинчина, где сказано, что поток требований является не чем иным, как дискретным случайным процессом, принимающим только целочисленные неотрицательные значения. Нужно, однако, отметить, что такой подход становится теперь уже недостаточным, так как в ряде важных задач необходимо изучать поток требований не только во времени, но и в пространстве. Отсюда возникает мысль о следующем определении.

Пусть задано множество элементарных событий X , а также σ -алгебра его измеримых подмножеств $\mathfrak{A}_X$. Случайным потоком $\eta(A)$ с фазовым пространством $(X, \mathfrak{A}_X)$ называется система случайных величин $\eta(A)$, определенных на элементах $A \in \mathfrak{A}_X$ и обладающих следующими двумя свойствами:

- 1) $\eta(A)$ — абсолютно аддитивная функция множества A ;
- 2) $\eta(A)$ принимает лишь неотрицательные целочисленные значения.

Для примера, если нас интересует изучение потока землетрясений, то указания только момента t наступления землетрясения недостаточно, необходимо дополнительно указать и координаты его эпицентра φ, θ, h — широту, долготу и глубину. Таким образом, в качестве множества X следует принять множество точек (t, φ, θ, h) , где $0 \leq \varphi \leq 2\pi$, $0 \leq \theta \leq \pi$, $0 < h < 6000$ км, $0 < t < \infty$.

Важным частным примером случайного потока является пуассоновский поток. В соответствии с предложенным общим подходом, пуассоновским потокам следует дать такое определение. Пусть $\eta(A)$ — абсолютно аддитивная, неотрицательная функция множества A , определенная на $A \in \mathfrak{A}_X$. Для любой системы $A_1, A_2, \dots, A_n$ попарно непересекающихся множеств величины $\eta(A_1), \eta(A_2), \dots, \eta(A_n)$ взаимно независимы и для каждого допустимого множества имеют место равенства

$$P\{\eta(A) = k\} = \frac{[\Lambda(A)]^k}{k!} e^{-\Lambda(A)}.$$

Не изменяя известных рассуждений монографии Хинчина, легко доказать, что поток $\eta(A)$ тогда и только тогда определяется равенствами

$$P\{\eta(A) = k\} = \frac{(\lambda |A|)^k}{k!} e^{-\lambda |A|}.$$

(простейший поток), где $\lambda > 0$ — постоянная, $|A|$ — мера множества A , когда выполнены условия стационарности, ординарности и отсутствия последействия.

Более того, многие теоремы, доказанные Хинчиным относительно строения потоков и помещенные в настоящей книге, легко переносятся и на приведенное здесь определение потока.

Предельная теорема, вскрывшая одну из существенных причин, в силу которой простейший поток может служить в широких границах хорошим приближением к истинному потоку требований, была предметом исследований ряда ученых. Прежде всего, как я упомянул в примечании к основному тексту (см. стр. 60), Г. А. Ососков обнаружил, что условие Хинчина является не только достаточным, но по сути дела и необходимым.

Позднее И. Б. Погожев заинтересовался вопросом быстроты сближения с пуассоновским распределением в зависимости от числа слагаемых потоков. Для случая одинаково распределенных независимых стационарных потоков он нашел первый член асимптотического разложения по степеням n^{-1} (работа не опубликована, докладывалась на семинаре по теории массового обслуживания, МГУ, 1961 г.). Более детальное исследование этого вопроса в тех же условиях под влиянием доклада И. Б. Погожева было проведено П. Франкеном. Б. И. Григелионис применил к доказательству предельной теоремы Хинчина-Ососкова методы классической теории суммирования. Случай суммирования нестационарных потоков изучен Б. И. Григелионисом. Им найдены необходимые и достаточные условия сближения суммарного потока с нестационарным пуассоновским; эти условия особенно просты в том случае, когда суммируются потоки типа процессов восстановления. Им же рассмотрен во всей полноте вопрос асимптотического разложения.

В системах с многоступенчатым обслуживанием, когда требование, после того как оно обслужено прибором первой

группы, поступает на прибор второй группы и т. д., интерес представляет не только поток входящих требований, но и те новые потоки, которые образуются требованиями, уже обслуженными приборами первой группы (или же получившими на них отказ), а также второй и последующими группами приборов. С такой ситуацией приходится иметь дело во многих реальных проблемах, в частности при расчете автоматических линий. К сожалению, это направление исследований находится совсем в зачаточном состоянии. Кроме результата Пальма, сформулированного и доказанного в монографии А. Я. Хинчина «Математические методы теории массового обслуживания» (стр. 107, § 28), можно отметить несколько статей Н. В. Яровицкого [1], [2]. Н. В. Яровицкий подверг систематическому изучению поток выходящих после обслуживания требований и ввел в рассмотрение односвязно зависимые потоки. Мы направляем читателя за определением потоков этого типа и их свойствами к указанным работам автора.

Во многих случаях первоначальный поток, проходя через ряд последовательных установок, постепенно теряет часть своих требований. Так обстоит дело на автоматической линии, когда после каждой операции происходит отбраковка обрабатываемых изделий. Подобная же картина наблюдается при исправлении опечаток несколькими корректорами. Поток опечаток при этом редееет. Перед исследователями неоднократно возникал вопрос относительно свойств таких редееющих потоков. Одно из предположений состояло в том, что при весьма общих условиях, касающихся процесса разряжения, поток требований должен подходить все ближе и ближе к семейству пуассоновских. Мы сформулируем сейчас теорему, полученную в этом направлении А. Реньи [1].

Пусть поток A подвергнут следующей операции: каждое требование потока остается с вероятностью p и выбрасывается из потока с вероятностью $q = 1 - p$; одновременно изменяется масштаб, причем за единицу масштаба времени принимается величина $1/q$. Обозначим эту операцию через T_p .

Если к потоку A с ограниченным последствием и конечной интенсивностью λ применить последовательность операций разряжения $T_{p_1}, T_{p_2}, \dots, T_{p_n}$ и при $n \rightarrow \infty$

$$p_1 p_2 \dots p_n \rightarrow 0,$$

то поток $T_{p_1}, T_{p_2}, \dots, T_{p_n}$ стремится к простейшему потоку интенсивности λ .

Обобщение этого результата было дано Ю. К. Беляевым.

Им же [1] был указан один тип потоков, представляющий интерес для теории массового обслуживания.

В последнее время значительное число работ было посвящено вопросам обслуживания неординарного потока, поскольку в реальных задачах приходится считаться с тем, что требования могут появляться не поодиночке, а группами. Изучением неординарных потоков без последствий систематически занимается П. И. Васильев.

Системы с потерями. Классическая проблема обслуживания с потерями за последние годы получила значительное развитие и в некоторых аспектах может считаться уже исчерпывающе разрешенной. Так, в примечаниях к тексту второй части монографии Хинчина было отмечено (стр. 84), что формулы Эрланга сохраняют свой вид в случае, когда обслуживается стационарный пуассоновский поток n приборами одинаковой производительности, а функция распределения длительности обслуживания произвольна с конечным математическим ожиданием. Метод доказательства, примененный Б. А. Севастьяновым [1], оказался особенно удачным. Выяснилось, как мы отчасти увидим позднее, что его рассуждения применимы ко многим задачам более сложной природы.

Другое направление обобщений формул Эрланга было развито Пальмом [1] и Такачем [2], [3]. Они рассмотрели случай, когда на обслуживающую систему поступает произвольный поток типа Пальма, а длительность обслуживания каждого требования (на каждом приборе) имеет распределение $H(x) = 1 - e^{-\mu x}$.

В работе [3] Л. Такач обобщил эту постановку задачи на следующую схему: требование остается в системе, если имеется хотя бы один свободный обслуживающий прибор или если число ожидающих требований не превосходит заданного числа m ; в противном случае требование покидает систему обслуживания.

А. Шахбазов рассмотрел случай обслуживания пуассоновского потока требований (с потерями) приборами разной производительности.

В последних работах В. Бенеша [1] имеются попытки изучить вероятность потери при обслуживании произвольного нестационарного потока.

Еще в 1918 г., вскоре после опубликования работы К. Эрланга, Энгсет обобщил результат Эрланга на случай обслуживания конечного множества потребителей, каждый из которых может предъявить требование в случайный момент времени. К этой задаче возвращались неоднократно, в частности при изучении простоев механизмов, обслуживаемых одним рабочим или бригадой. Мы укажем на большое исследование Ж. Козна [1]. В теоретическом и прикладном отношении исключительно интересно при этом учесть то обстоятельство, что требование, получившее отказ, вновь и вновь предъявляет свои претензии к системе обслуживания.

Полезно указать еще на два направления обобщений постановок вопросов об обслуживании требований с потерями. Первое из них состоит в том, что на систему приборов поступают несколько потоков, причем потоки с большими номерами обладают преимуществами по сравнению с предшествующими. Это преимущество состоит в том, что требование с большим номером, поступившее в систему и заставшее все приборы занятыми, вытесняет требование с минимальным номером (это последнее может теряться или становиться в очередь) и может быть потеряно лишь в том случае, когда все приборы заняты обслуживанием требований либо того же номера, либо более высоких номеров.

В частности, именно с такой постановкой задачи приходится иметь дело, когда сами обслуживающие приборы могут выходить из рабочего состояния. Тогда следует учитывать сразу два потока: поток требований и поток поломок, причем поток поломок обладает преимуществом. Недавно в работе Т. П. Марьяновича [1] методом Севастьянова была изучена следующая задача: простейший поток интенсивности λ поступает на систему, состоящую из n одинаковых приборов. Каждый прибор независимо от других может выйти из рабочего состояния в любой момент времени в период, когда он занят обслуживанием. Период непрерывной работы прибора — случайная величина с функцией распределения $K(x)$. Время восстановления прибора —

случайная величина с функцией распределения $G(x)$. Длительность обслуживания требования — случайная величина с функцией распределения $H(x)$. Предполагается, что $H(x)$ имеет конечный первый момент, а также хотя бы одно из распределений $K(x)$ или $G(x)$ имеет конечный первый момент. В этих предположениях доказано существование пределов $p_{ij} = \lim_{t \rightarrow \infty} P_{ij}(t)$, где $P_{ij}(t)$ — вероятность того, что в момент t обслуживанием заняты i приборов, а j приборов находятся в ремонте ($i, j = 0, 1, 2, \dots, n; i + j \leq n$). Требование теряется, если оно поступило в момент, когда все приборы заняты (или находятся в ремонте), а также тогда, когда обслуживающий прибор вышел из рабочего состояния. В последнем случае требование теряется даже тогда, когда в системе имеются свободные приборы.

Только что описанная задача была обобщена Т. П. Марьяновичем на тот случай, когда в системе имеются, помимо приборов работающих, также приборы, находящиеся в резерве разного порядка [2]. Понятно, что эта задача имеет самое непосредственное отношение к вопросам теории надежности.

Второе направление обобщений касается выбора оптимальных режимов обслуживания.

Имеется n приборов разной производительности: среднее время обслуживания требования для них соответственно равно $a_1, a_2, \dots, a_n$. Приборы расположены в определенном порядке, и требования потока поступают на свободный прибор с наименьшим номером в только что определенном порядке. Требование, заставшее все приборы в работе, теряется. Спрашивается, как надо перенумеровать приборы, чтобы вероятность потери была минимальной? Ответ таков: приборы следует нумеровать так, чтобы с ростом номера уменьшалась их производительность.

Следующая задача в том же круге идей была рассмотрена Г. П. Климовым. Имеется группа однотипных приборов и заданного объема накопитель, в котором требования могут ожидать освобождения приборов. Требования, попавшие в систему, когда накопитель полон, теряются. На систему обслуживания поступает пуассоновский поток с параметром $\lambda(t)$. Спрашивается, как выбрать $\lambda(t)$, чтобы итог работы системы за время $(0, T)$ был максимален. Итог оцени-

вается функционалом $I(\lambda(t), \mu)$. Длительность обслуживания случайна и имеет показательное распределение: $1 - e^{-\mu t}$.

Таким итогом работы могут быть общая производительность системы, количество потерь и пр.

Оказывается, что экстремум достигается на классе ступенчатых функций $\lambda(t)$. В частности, для получения максимальной производительности оборудования следует выбрать сначала $\lambda(t)$ возможно большим, а затем сделать минимально допустимым. Этим достигается быстрое заполнение накопителя. Наличие требований в накопителе обеспечивает бесперебойную работу механизмов системы.

Системы с ожиданием. Пожалуй, максимальное число работ, относящихся к теории массового обслуживания, посвящено изучению образования очередей в системах с ожиданием. Распределение длительности ожидания, условия образования увеличивающейся очереди, распределение длительности непрерывной занятости прибора — вот примерный перечень вопросов, которым было уделено много внимания. Особенно интересовал исследователей случай одного обслуживающего прибора. Мы знаем, что для пуассоновского потока и произвольного распределения длительности обслуживания общие результаты были получены А. Я. Хинчиным. Изящные результаты были получены здесь также Линдли, Л. Такачем, А. Шахбазовым [1] и др. Линдли получил интегральное уравнение для распределения длительности ожидания начала обслуживания в случае одного прибора для произвольного пальмовского потока и произвольного распределения длительности обслуживания. Это интегральное уравнение дало ему возможность заключить, когда очередь будет расти неограничено и когда распределение величины очереди стремится к стационарному состоянию.

Обслуживание одним прибором неординарного потока было предметом исследований Гэйвера [1], Миллера [1] и Шахбазова [1]. А. Шахбазов ввел усложнение, представляющее интерес для теории надежности, а именно: он предположил дополнительно, что прибор, начиная обслуживание требования без предварительного ожидания его освобождения, требует времени на «разогрев».

Киффер и Вольфовиц [1], [2] исследовали обслуживание произвольного пальмовского потока при условии произвольного распределения длительности обслуживания n одинако-

выми приборами. Ими составлены интегральные уравнения задачи, получено условие существования стационарного решения.

Известно, что если интенсивность поступления требований в систему равна средней производительности системы, то длина очереди со временем возрастает неограничено (за исключением тривиального случая регулярного поступления требований, равного длительности обслуживания). Возникает вопрос изучения асимптотического поведения различных характеристик обслуживания. Этот вопрос интересовал многих исследователей. Я приведу здесь результаты И. Н. Коваленко.

Пусть λ — интенсивность входящего потока, τ — средняя длительность обслуживания одного требования, число обслуживающих приборов равно единице и $\rho = \lambda\tau = 1 - \varepsilon$, где $\varepsilon > 0$. Если $\varepsilon \rightarrow 0$, то средняя длительность времени ожидания начала обслуживания будет расти, имея порядок $1/\varepsilon$. Функция распределения длительности ожидания y подчинена такому предельному распределению

$$P\{\varepsilon y > x\} \rightarrow e^{-\frac{\sigma^2 x}{\sigma^2 + \tau^2}} \quad (\varepsilon \rightarrow 0),$$

где σ^2 — дисперсия длительности обслуживания.

Точно так же, если дисперсия времени обслуживания конечна, то для распределения числа v требований в системе имеет место предельная теорема: при $\varepsilon \rightarrow 0$

$$P\{\varepsilon v > x\} \rightarrow e^{-\frac{\sigma^2 x}{\sigma^2 + \tau^2}}.$$

Большое внимание уделялось последние годы изучению обслуживания с очередью при наличии преимущественных требований. Система обслуживания может быть описана таким образом: имеется k классов требований, каждый из которых образует простейший поток; потоки между собой независимы. Внутри каждого класса обслуживание производится в порядке строгой очередности. Различаются две схемы: обслуживание без прерывания (английский термин — head of the line) и обслуживание с прерыванием (preemptive priority). Предположим, что в некоторый момент времени происходит обслуживание требований i -го класса и поступает требование класса j ($j > i$); для первой схемы вновь

поступившее требование ожидает окончания обслуживания низшего класса, но зато немедленно занимает первую освободившуюся линию, если только перед ним не было требований того же или более высокого класса. Вторая схема в свою очередь подразделяется на две категории: в первой из них учитывается время, уже потраченное на его обслуживание (*preemptive resume*), во второй (*preemptive repeat*) это время теряется и обслуживание начинается сначала. Отметим здесь лишь работу Миллера [2], в которой указаны предшествующие исследования, а также для случая $k=2$ даны такие характеристики: производящая функция стационарных вероятностей числа преимущественных и не преимущественных требований, находящихся в пункте обслуживания, преобразования Лапласа для функций распределения времени обслуживания и периода занятости, а также для числа требований, обслуженных за один период занятости.

Как частный случай указанной схемы, можно рассмотреть такую задачу, имеющую очевидный практический интерес: изучить систему обслуживания с очередью при дополнительном предложении, что обслуживающие приборы могут выходить из рабочего состояния и их восстановление требует, вообще говоря, случайного времени. Здесь мы виювь имеем два потока: поток требований и поток поломок, причем поток поломок пользуется преимуществом. Иногда полезно учитывать также необходимость профилактических мероприятий.

В ряде телефонных вопросов приходится встречаться с такими постановками задач, когда один поток обслуживается по схеме ожидания, а другой по схеме потерь.

Для ряда практически важных задач интересно рассмотрение задачи обслуживания с очередью при дополнительном вмешательстве регулирующей силы. Если иметь в виду интересы работников морских портов, то задача может быть поставлена так: имеется группа близких портов, в которых имеются соответственно $n_1, n_2, \dots, n_k$ причалов. В том случае, когда какой-либо порт перегружен и имеется опасность возникновения очереди, диспетчер отдает распоряжение о подходе очередного судна в тот порт, в котором имеются свободные причалы или же причалы должны освободиться в ближайшее время. Само собой разумеется, что

в экономическом отношении эта задача представляет значительный интерес. Кстати следует сказать, что представляет интерес изучение различных характеристик экономической эффективности использования портов. Так, некоторые экономисты ставят себе в особую заслугу то, что удается добиться рекордного времени использования причалов. При этом никогда не отмечается, как долго ждут погрузки или разгрузки суда, как велики их непроизводительные стоянки в ожидании освобождения причала.

Обслуживание с ограничениями. Как ни важны в прикладном отношении и как ни интересны в теоретическом отношении задачи, возникающие при изучении обслуживания с очередью и с потерями, в настоящее время, по-видимому, большее значение приобретает объединяющая обе эти задачи схема.

Мы называем эту схему обслуживанием с ограничениями. Она включает в себя большое число разнообразных и практически интересных вопросов.

Все возникающие здесь задачи так или иначе связаны с тем, что требование, попавшее в систему обслуживания и заставшее все приборы занятыми, не обязательно теряется, а оставшись в очереди, не обязательно доводит ожидание до начала обслуживания.

Как правило, требование ограничено в своих возможностях ожидания и даже пребывания в системе (ожидание плюс обслуживание). Так, для примера, человек, пришедший на междугороднюю телефонную станцию, готов уделить некоторое время на ожидание начала разговора, однако он не может ожидать сколь угодно долго, так как либо занят другим делом, либо же слишком длительное ожидание приведет к тому, что надобность в разговоре отпадет. Точно так же самолет, вставший в очередь на посадку, ограничен в своих возможностях ожидания многими обстоятельствами: количеством горючего в баках, состоянием экипажа и пр.

Последние годы многие авторы систематически подвергали изучению такую схему, которая учитывает возможность очереди ограниченного размера (всякое требование, заставшее в системе m ожидающих, теряется). Далее, рассматривался случай, когда требование, заставшее все приборы занятыми, остается в системе лишь с вероятностью, зависящей от

длины очереди. Исходя из разных прикладных задач, почти одновременно ряд авторов пришли к постановке близких задач теории обслуживания в таких условиях:

1) требование, поступившее в систему обслуживания, ожидает начала обслуживания в течение времени τ (вообще говоря, случайное);

2) требование остается в системе обслуживания не более чем время τ (даже в том случае, если его начали обслуживать, но длительность обслуживания превышает τ).

Понятно, что обе эти постановки задач включают в себя как крайние случаи схемы обслуживания с потерями и с очередью.

Случай $n=1$ был детально исследован И. Н. Коваленко [1]. В этой работе он рассмотрел несколько более широкую схему, чем мы только что наметили. В ней учтено то обстоятельство, что длительность пребывания зависит от того, сколько времени требование ожидало начала обслуживания.

Теория обслуживания с ограничениями находится еще в начальной стадии, и несомненно, ближайшее будущее принесет многочисленные полезные результаты, которые не только дадут возможность решить те или иные частные прикладные задачи, но и создать общую теорию, способную охватить многочисленные постановки вопросов единым подходом, единой точкой зрения.

Обратим внимание на то, что подавляющее число задач теории массового обслуживания приводит к необходимости рассмотрения процессов, отличных от марковских. Нередко удается тем или иным способом (например, путем введения дополнительной переменной) превратить эти процессы в марковские и тем самым получить возможность использовать хорошо разработанный аналитический аппарат теории марковских процессов. На этом пути получено большое число ценных результатов.

Отметим, что в этой небольшой обзорной статье не было возможности даже только упомянуть все наиболее интересные работы. Довольно полная библиография (до 1957 г.) составлена Дойгом [1]. Она может служить хорошим дополнением к библиографии, имеющейся в уже упомянутых книгах Сиски и Саати.

ЛИТЕРАТУРА

- A. T. Bharucha—Reid [1], Elements of Markov processes and the applications, McGraw Hill, 1960.
- Беляев Ю. К. [1], Линейчатые марковские процессы и их приложение к задачам теории надежности.
- V. E. Benes [1], General Stochastic Processes in Traffic Systems with One Server, Bell System. Techn. J., 1960, 127—159.
- G. A. Bodino, F. Bionbilla [1], Teoria delle code, Milano—Vares, 1959, 1—219.
- D. P. Gaver [1], Imbedded Markov chain analysis of a Waiting line Processes in continuous time, Ann. Math. Statistics, v. 30, 1959, 698—720.
- A. Doig [1], A bibliography on the theory of queues, Biometrika 44, № 3—4, 1957, 490—514.
- Kiefer and Wolfowitz [1], On the theory of queues with many servers, Trans. Amer. Math. Soc. 78, 1955, 1—18.
- [2] On the characteristics of the general queuing process with applications to random walk, Ann. Math. Statistics 27, № 1, 1956 г. 147—161.
- Коваленко И. Н. [1], Некоторые задачи массового обслуживания с ограничением, Теория вероят. и ее применения, т. 6, № 1, 1961, 222—228.
- Cox and Walter L. Smith [1], Queues, Methuen & Co, London, 1961.
- Cohen J. W. [1], The generalized Engset Formulae, Philips Telecommunication Rev., 18, № 4, 1957, 158—170.
- Марьянович Т. П. [1], Обобщение формул Эрланга на случай, когда приборы могут выходить из строя и восстанавливаться, Укр. матем. ж., т. 12, № 3, 1960, 279—286.
- [2]. Надійність системи зі змішаним резервом, Доповіді УРСР, № 8, 1961, 994—997.
- Rupert G. Miller, [1] A contribution to the theory of Bulk Queues, J. Roy. Statist. Soc. B, v. 21, № 2, 1959, 320—337.
- [2] Priority Queues, Ann. Math. Statistics, 31, № 1, 1960, 86—106.
- Palm C., [1], Intensitätsschwankungen im Fernspreverkehr, Ericsson Techn., 44, 1943, 1—189.
- Rényi A., [1], A Poisson—folymat egy jellemzése, Tr. Matem. ин-та Венгрин, v. 1, fasc. 4. 1956, 519—527.
- J. Riordan, [1], Stochastic Service Systems, Wiley and Sons, 1962, 1—139.
- Saaty T. L., [1] Résumé of useful formulas in Queuing Theory, Operat. Res., v. 5, 1957, 162—187.
- [2] Mathematical methods of operations research, McGraw-Hill, 1959, 331—374.
- [3] Elements of Queuing theory with applications, McGraw-Hill, 1961, 1—423.
- R. Syski, [1], Introduction to congestion theory in telephone systems, Oliver and Boyd, Edinburgh and London, 1960, 1—742.

- Севастьянов Б. А. [1], Эргодическая теорема для марковских процессов и ее приложение к телефонным системам с отказами; Теория вероят. и ее применен., т. 2, вып. 1, 1957, 106—116.
- Takács L., [1] Некоторые вероятностные задачи в телефонии, Математика (сб. переводов), 4:6, 1960, 93—144.
- [2] Introduction to the theory of queues, Oxford University Press, 1962, 1—268.
- [3] On the generalization of Erlang's formula, Acta Math. Acad. scient. Hung., 7, 1956, 419—433.
- [4] On a combined waiting time and loss problem concerning telephone traffic, Ann. Univ. Sc., t. 1, Budapest, 1958, 73—82.
- Шабазов А., [1] Обслуживание неординарного потока; Докл. АН СССР, т. 145, № 2.
- Яровицкий Н. В., [1] Протвіхідний потік одиолінійної системи обслуговування з втратами, Доповіді АН УРСР, № 10, 1961, 1251—1254.
- [2] О некоторых свойствах односвязно-зависимых потоков, Укр. матем. ж., т. 14, № 2, 1962, 170—179.
-

Александр Яковлевич Хинчин

**Работы по математической теории
массового обслуживания**

М., Физматгиз, 1963 г., 236 стр.

Редактор Морозова И. Е.

Техн редактор Михлин Э. И.

Корректор Володьева И. А.

Сдано в набор 30/VI 1962 г. Подписано к печати
16/I 1963 г. Бумага $84 \times 108^{1/32}$. Физ. печ. л.
7,375, Условн. печ. л. 12,92. Уч.-изд. л. 11,34.
Тираж 15 000 экз. Т-10925, Цена книги 72 коп.
Заказ № 3121.

Государственное издательство
физико-математической литературы,
Москва, В-71, Ленинский проспект, 15.

Первая Образцовая типография
имени А. А. Жданова
Московского городского совнархоза,
Москва, Ж-54, Валовая, 28.

Отпечатано с матриц в Гос. тип. «Вайздас»,
Вильнюс, Страдылио 1. Заказ № 238.